The Grammar of Immersion: A Social Semiotic Study of Nonfiction Cinematic Virtual Reality

Phillip Doyle Dip., B.A., M.Sc.

Thesis presented for the award of PhD, August 2023

School of Communications, Dublin City University

Supervisors: Dr. Declan Tuite & Prof. Patrick Brereton

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of Doctor of Philosophy is entirely my own work, and that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge breach any law of copyright, and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Signed: __Phillip Dyer__ *(Candidate) ID No.: 15211932 Date: __18/08/23__*

Acknowledgements

I would like to express my gratitude to everyone who helped me throughout the course of my PhD studies.

First, a major thanks to my supervisors Dr. Declan Tuite and Prof. Patrick Brereton for their balance of personal support, encouragement and incisive feedback. As well guiding me on my own PhD path, Dr. Tuite and Prof. Brereton gave me valuable insight into the practice of postgraduate supervision, which I will try to take with me into the next phase of my career.

I would also like to extend my appreciation to those involved in the assessment of my thesis: Dr. Miriam Judge and Prof. Paul McKevitt who examined my thesis and Dr. Peter Admirand who acted as chair. I will remember the viva examination as a highlight of my academic career.

On a personal note, I would like to thank my family for their support over the years. Thank you to my parents Maggie and Des for helping me to get started on my academic journey. Special thanks to my partner Amy for encouraging me to start my PhD - and for patiently waiting for me to finish it. Finally, thanks to Jack and Lucy who motivate me to do my best in life, of which this PhD is a significant part.

Table of Contents

List of Tables

List of figures

Glossary of Technical Terms and Abbreviations

360° image	The colloquial term used to describe the CVR image surrounding the spectator in all directions.
ASMR	'Autonomous sensory meridian response'. A physiological response to audio- visual material
CAQDAS	Computer Aided Qualitative Discourse Analysis Software
CRU	'Conjunctively relatable unit'. A generic unit used to track logical meanings across multimodal entities.
CVR	Shorthand for 'cinematic virtual reality', the term used to denote linear video content displayed in a VR headset.
Fisheye	The optical effect of wide-angle lenses including CVR camera lenses. When the 360° image is displayed on flat screens, depth is enhanced and objects in the centre of the screen look far away.
FGM	'Female Genital Mutilation' A ritualistic operation performed on young women, intended to remove the capacity for sexual feelings in preparation for adult life and marriage.
GoPro	A brand of durable, waterproof action cameras. The brand has become genericized to refer to the class of cameras.
HMD	'Head mounted display'. The headset worn by the CVR viewer which allows them to rotate their view around the 360° image.
Immersion	The technical effect of wearing the headset which gives the spectator access to the full 360° image while blocking their immediate (actual) surroundings
L.A.P	'Learning Assistance Programme'. A training programme instituted for young women in Sierra Leone to serve rural communities and schools underrepresented by female staff
Mediography	A database of multimedia artefacts centred on a subject or theme
POV	'Point of view'. The kind of view enabled by the placement of the camera.
Presence	The psychological effect of being transported to the 'place' represented in the 360° image.
SEF	A sensory ethnographic film
SEL	'Sensory ethnography Laboratory'. An MIT research laboratory that privileges (technologically enabled) sensory knowledge, over disembodied knowledge.

SFL	Systemic Functional Linguistics. The linguistic framework used as the basis of the study.
SFMDA	Systemic Functional Multimodal Discourse Analysis. The multimodal extension of the SFL framework, used as the multimodal basis of this study.
Spectator	The individual viewer of CVR texts.
Spherical image	A visual metaphor describing the 360° image. The 'sphere' is the CVR video imagined to envelope the spectator.

Guide To the Texts

A brief overview of each of the films' production details is included below followed with a full list of shots from each of the films, including timecode. For brevity, I abbreviate the film titles throughout the document when referring to specific shots, for example 'M#1' for *Mamie's Dream*, *shot 1*, and 'C#5' for *Congo VR, shot 5*. I will include timecodes in-text, when discussing a specific moment within a shot, and omit it when discussing actions that span entire shots.
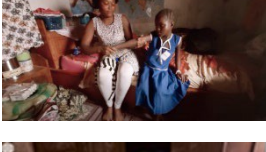
Film Title: Mamie's Dream
Director: Mary Matheson
Producer: Plan International
UK **Link to film:**
bit.ly/3myYUMr **Running time:** 8:02
Format: Stereo sound / monoscopic 360° VR
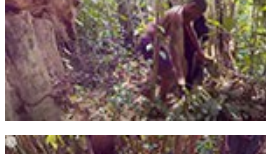
Film Title: Congo VR: A Troubled Past
Director: Phil Harper
Producer: BBC & Zillah Watson
Link to film: https://bit.ly/397mwEI
Running time: 10:01
Format: Stereo sound / stereoscopic 360° VR

Mamie's Dream. Shot Listing

	Title / Timecode		**Title / Timecode**
	0 TITLES **00:00**		10 ADULT LEARNERS **3:44**
	1 IN BOAT **00:07**		11 MAMIE ASSISTING **4:06**
	2 ON BRIDGE **00:37**		12 MAMIE TEACHING **4:20**
	3 KIDS PLAYING FOOTBALL **00:54**		13 OPEN AREA **4:40**
	4 UNDER TREE IN YARD **1:17**		14 IN CHURCH **4:57**
	5 STANDING IN DOORWAY **1:37**		15 15 HOLDING HANDS **5:13**
	6 SCHOOLROOM 1 **1:54**		16 GIRLS PLAYING **5:34**
	7 SITTING ON BED **2:19**		17 SERIOUS CHAT **5:56**
	8 IN MARKET **2:47**		18 DANCING CLAPPING **6:27**
	9 SITTING UNDER SHELTER **3:21**		19 RIVER **6:49**

Congo VR: A Troubled Past. Shot Listing

	Title / Timecode		Title / Timecode
	1 DRONE OVER DELTA **0:14** *(titles overlaid)*		**11 KINSHASA TRAFFIC** **2:34**
	2 DRONE FORESTRY **0:26**		**12 DRONE ROOFTOPS** **2:49**
	3 DRONE SANDY LANDSCAPE **0:39**		**13 JUNIOR** **3:01**
	4 PRESENTER IN BOAT **0:53**		**14 YARD PRINCESS MAMICHO** **3:16**
	5 DRONE RIVER TITLE **1:06**		**15 DANCELFOOR** **3:40**
	6 DRONE RIVER RAPIDS **1:22**		**16 DRONE STATIC FORESTRY** **4:26**
	7 TRAIN STATION **1:39**		**17 VILLAGE TRIBE** **4:44**
	8 TRAIN INTERIOR WITH A.L. **1:49**		**18 FROEST HUNT** **5:03**
	9 TRAIN INT. WINDOW **2:01**		**19 FOREST HUNT #2** **5:30**
	10 TRAINSTATION #2 **2:14**		**20 FOREST HUNT #3** **5:49**

Abstract

The Grammar of Immersion: A Social Semiotic Study of Nonfiction Cinematic Virtual Reality.

Phillip Doyle

Cinematic virtual reality (CVR) is an audio-visual form viewed in a virtual reality headset. Its novelty lies in the way it immerses its audience in highly realistic 360° visual representations. Being camera-based, CVR facilitates many of the practices of conventional filmmaking but fundamentally alters them through its lack of a rectangular frame. As such, CVR has garnered scholarly attention as a 'frameless' storytelling medium yet to develop its own language. The form has gained traction with producers of nonfiction who recognize CVR's capacity to transport audiences to remote social worlds, leading to claims that equate CVR's immersion with a social and emotional response to its filmed subjects. A strand of CVR scholarship has emerged, grounding nonfiction CVR theoretically and critiquing such deterministic claims. Broadly speaking, these parallel strands of inquiry point to a common concern with CVR's semiotics; as the meaning potential of the 360° format, and the social aspects of its use in documenting reality. Currently however, there appears to be a lack of systematic analyses that foreground CVR's semiotics.

This study addresses this gap by using social semiotic methods to complement these threads of inquiry, subsuming them into a holistic account of CVR's semantics. Utilizing systemic functional methods, multimodal discourse analyses were performed on nonfiction CVR texts addressing core research objectives. The first objective is the systematic description of CVR as a semiotic technology, and the configuring of discourse through its novel 360° modality. The CVR spectator is described for their role in the real-time construction of low-level meanings. Higher-level concepts further characterize CVR texts as technologically enabled, virtual sites of social discourse. The second research objective concerns clarifying the implications of CVR for nonfiction practitioners. Nonfiction discourse is conceptualized as the negotiation of semiotic autonomy, independence, and control, between viewing spectator, filmed subject, and CVR author respectively. The third objective concerns the development of an analytical approach tailored specifically for CVR. Extant systems from image, text, film, and action analyses are reflexively applied, appraised, and adapted for use in the study of CVR and new frames are presented to cater for the 360° modality.

The findings show CVR to be an inherently logical, contextualizing form, where the spectator has a degree of sense-making autonomy in the construction of representational and social meanings. This semantic autonomy is found to camouflage the deeper textual constructions in what appear as 'reality experiences'. The repercussions for the CVR producer are the indeterminacy of meanings which are 'at risk' in particular ways when conventional framing methods cannot be utilized, and when the spectator is given reflexive agency to make meaningful connections across the 360° image. Systemic functional analytical methods prove flexible enough to be applied to the texts, and open enough for the study to present additional systems and frames for a more fulsome approach to the analysis of CVR.

1. Introduction

Cinematic virtual reality (CVR) is an audio-visual form viewed in a virtual reality headset. Its novelty lies in the way it immerses its audience in highly realistic 360° visual representations. CVR is currently studied for the ways that its users engage with its novel image modality. From a cinematic perspective, it is studied for its limitations compared to conventional film as well as its unique potential as a storytelling medium. Nonfiction CVR is studied for the ethical and epistemological implications of an immersed spectator brought into contact with the social world in seemingly unmediated experiences. These disparate scholarly domains essentially study *meaning* in CVR, as the production of meaning and as the critique of what constitutes meaningful experiences of reality. As such there is an opportunity to incorporate these endeavours into a singular systematic account of CVR from a textual, semantic perspective.

At present there appears to be a lack of systematic accounts of CVR that foreground the underlying semiotics of its novel modality. This study aims to bring together the practical and social aspects of CVR, by conducting a social semiotic analysis of nonfiction CVR texts. This will identify the ways that CVR's technological makeup impacts on the construction of its discourse, while maintaining a focus on the social and ideological aspects of its production as it is used to document the social world.

This introductory chapter provides an overview of the study by adding detail to the context and background of the study, the problem that the study addresses, its aims and conceptual framing, and its significance and limitations. The chapter is concluded with a synopsis of each of the thesis' chapters.

1.1   Background

CVR can be described as a hybrid technology as it fuses elements of linear video with VR display mechanisms. The VR headset functions to erase itself by presenting the spectator with an all-encompassing 360° image of the scene captured by the camera. CVR is thus lacking the rectangular frame that has become synonymous with visual cinematic meanings. Multiple studies have aimed to reimagine CVR as a frameless storytelling medium (Bucher, 2017; Dooley, 2017a; Tricart, 2017; Elmezeny, Edenhofer and Wimmer, 2018; Zhang and Weber, 2021). Others have focused on specific challenges presented by CVR's frameless-ness, such as the perennial issue of controlling of the spectator's attention (Brown et al., 2016a; Brillhart, 2018b). Where conventional cinema presents framed images in rapid succession, CVR presents longer-duration omnidirectional images where the spectator can survey the contents of the shot at their discretion. This makes linear storytelling difficult, where to rotate one's view at the

wrong time will result in missed plot-points. To address this, Brown et al. (2016a) identified the kinds of cues that best attract attention, and Brillhart (2018) and Kjaer et al. (2017) have posited ways to maintain their attention across multiple shots. Conventional cinematic shot types have been interrogated for their use in CVR, such as the ubiquitous close-up shot, which becomes largely redundant in CVR (Dooley, 2020) as *close-up* necessarily means *close-to*. Passmore et al. (2017) take a more holistic approach to CVR 'literacy' posing the question of whether other paradigms are more feasible, where the spectator is not predisposed to linear engagement with CVR texts.

CVR's technological makeup has led to a strand of inquiry relating to its uses in the production of nonfiction. CVR functions largely through erasure: at the time of image capture, the CVR producer must either vacate the scene, or become part of the omnidirectional image, and at the time of viewing, the headset erases its own presence and the spectator's physical surroundings. It is as such a 'transparent' media form (Bolter and Grusin, 1999) that purports to afford immediacy between the spectator and a social reality. Deterministic claims have been made in this regard for CVR's capacity to automatically engender empathy with the filmed subject (Milk, 2016), when there is no visible mediating entity to get in the way. Conversely, CVR has been critiqued for the invisibility of the filmmaker from an ethical standpoint for its potential for propagandising, where the illusion of unmediated intimacy can be a subterfuge for invisible messaging (Kool, 2016a). In this regard there is a recycling of older claims regarding transparent documentary forms such as the observational movement of the 1960s and 70s, in which the invisibility of the filmmaker was rejected as a conceit with dangerous ideological consequences (Geiger, 2008; Waugh, 2011). Rose describes CVR's 'empathy machine' trope as simply recycling these older 'illusionistic' observational claims, ultimately hindering the form (Rose, 2016). Nash (2017) and Gregory (2016) approach nonfiction CVR from similar perspectives, of *media witnessing*, asking whether CVR is in fact commensurate with empathic encounters, and if empathy should be held as the ideal standard when the goal is to foster constructive audience engagements, rather than a purely emotional response (ibid.).

Essentially, these broad research endeavours relate in different ways to the envisioning of CVR's semiotics. Studies of CVR as a storytelling medium are essentially engaged in filmic meaning making resources, of how shots mean something specific in framed films but mean something else when the frame is removed. Critical studies of nonfiction CVR are engaging with higher-order semiotics and the construction of meaning that is both socially situated and technologically mediated. Currently there does not appear to be any studies of CVR that address these discrete areas of inquiry, holistically through a systematic account of CVR's semiotics.

1.2 Aims, Objectives, and Research Question

By adopting a social semiotic approach, this study aims to address CVR on practical and theoretical levels, as functional meaning-making resources which are necessarily employed in social and ideological contexts. As CVR is a relatively new form that has yet to be systematised from a semiotic standpoint, the study also reflexively contributes back into and extends multimodal systems and methodologies. These aims are reflected in the study's principal research objectives:

1. To describe CVR as a semiotic technology, in systematic terms.
2. To identify the implications for CVR's use in nonfiction
3. To establish a set of systems suitable for the analysis of CVR

The first research objective entails CVR's fundamental differences in the ways that the 360° image constrains and enables the flow of discourse. It also requires an account of how discourse is both cohesive and meaningful when the 360° image is juxtaposed with voice, music, and graphics. Meeting this objective provides a grounding for the subsequent objective where CVR's functional resources are contextualised for their use in nonfiction. The third objective is reflexively embedded in the first two, where CVR-specific analytical tools emerge in the process of doing the study.

By pursuing these aims and objectives in this way, the study assumes several factors that relate to the overarching social semiotic paradigm. First, nonfiction CVR in its current form is approached with a degree of scepticism, as I align myself with those who see nonfiction CVR as a technological spectacle that is at times incongruent with its aims. While I recognise the impetus behind the hyperbolic claims made on CVR's behalf, essentially to push the media form into the public consciousness, these claims pervade the resulting discourse and thus require systematic critique.

Another assumption relates to CVR practices generally. In the broad social semiotic frame, language-based practices can be considered to morph from framed to frameless production, where CVR producers adapt their previous understandings and aesthetics in response to the situations of CVR filming. Essentially, CVR producers are faced with a novel representational form, and strive to make sense of it based on extant practices.

This in turn raises an issue of technological determinism and the arguments for and against conceptualising technologies as *things* with agency in their own right. I take a middle position in this regard, following Gibson (1979) for whom technologies present 'affordances' (ibid), and a potentiality that is reflexively encountered by those who engage with the tools at hand. For

CVR producers it is the underlying impression of what this technology can do for them, that previous technologies couldn't.

It is from these positions that I take my stance on the evolution of CVR practices and choice of methodological framework for this study: that language, including visual language evolves through contexts that are both immediate and cultural, and can be understood in a structural fashion. As an educator in film and new media, I am naturally predisposed towards the systematising of meaning making. It is for these reasons that I have adopted a social semiotic approach to analysing CVR.

I envisage that the study will benefit analysts and practitioners engaging with CVR. In the case of those studying CVR, this thesis adds nuance in understanding the way meanings are encoded and decoded in CVR texts. At a high level, the study yields frames for considering CVR as a semiotic technology and a site of discourse production; at lower levels, several augmentations and additions to the methodologies are presented that are useful in further studies of CVR. Where the producer is concerned, the study contributes theoretically to debates about technological representations of the social world. From a practical perspective, the study interrogates the practical, semantic implications for documenting the social world in a manner that includes 360° images.

The limitations of the study relate to its scope. By studying the multi-level semantics of the form at a granular level, it is necessary to compensate in other ways. One trade-off is the small sample size required for depth, which cannot capture the full variety of nonfiction CVR as it exists today. This study looks at two different kinds of nonfiction texts, but there will no doubt be exceptions that can be claimed to the observations I make. I accept these limitations as necessary for a study of this kind, that necessarily compromises breadth for depth.

1.3 Layout of The Thesis

1.3.1   Chapter 2: Review of the Literature

The chapter is divided into three parts. First CVR is considered for the ways it can be conceptualised as an emerging technology, with remediation used as a theoretical frame. The form is then considered as a hybrid, as the confluence of other remediated forms (VR, film, theatre and photography for example). Finally, CVR is explored as a means of producing nonfiction texts.

The aim of the chapter is to identify relevant scholarship in the area and also to characterise nonfiction CVR in as much as is possible, prior to analysis. As part of this, I also present a dramaturgical frame used to describe nonfiction, which I bring to bear on CVR, in later chapters.

1.3.2   Chapter 3: The Systemic Functional Frame

Chapter three outlines the underlying paradigms on which the study was built. I describe the social semiotic underpinnings of the study, from the foundational *Systemic Functional Linguistics* to its outgrowths *Systemic Functional Multimodal Discourse Analysis*. The chapter provides the necessary theoretical grounding for my research methods, by establishing the core tenets of SFMDA on which I layer my own CVR specific augmentations. I will present a catalogue of analytical resources from language-based, action, image, and filmic texts for necessary context.

1.3.3   Chapter 4: Research Methods

Chapter four presents the methods used in the study. I begin with a rationale for choosing a small sample case study approach and outline how I consider this to be the most appropriate to the study's objectives. I then describe my approach to data collection, and the purposive sampling methods used to find the two texts used in the study. I discuss my reasons for choosing conventional qualitative analysis software over a multimodal analysis programme, and my reasons for specifically choosing MAXQDA to house the study. I describe the transcription process in terms of what I chose to transcribe to fully capture CVR's modality, why I chose it, and how I went about transcribing the data. Following this I present a round-up of the coding strategy used, comprising a system of SFMDA resources that were to be probed in analysis. To complete the chapter, I propose a system of CVR's 'grammatical' rank, and the grammatical units necessary in describing CVR systematically.

1.3.4   Chapter 5: Construing Experience

This chapter presents findings relating to analysis of the ideational metafunction, as the

configuration of representational meaning. It is structured in two parts: as a visual analyses of the texts' 360° images, and then as cross-modal analyses. Using the visual rank system proposed in chapter 4, I apply the principles of visual grammar to participants, processes, and circumstances across the 360° image. High-level conceptual frames are then used to establish the kinds of meanings that emerge between the image and the verbal modes.

The findings are framed for the ways the spectator engages with representations, and the kind of semiotic autonomy CVR affords. The producers are discussed for the kinds of semiotic risks attached to representing the social world with CVR.

1.3.5 Chapter 6: Interpersonal Relations

This chapter presents findings related to the CVR spectator's interpersonal construction through their perceived engagements with social actors in the texts. As with the previous chapter, this chapter is structured in two halves, first addressing the texts for visual interpersonal meaning systems and then for multimodal interpersonal meanings.

I present Adam Kendon's socio-spatial theories as conducive to the analysis of CVR's socio-spatial discourse configurations. Kendon's systems are used thereafter in the chapter as a means of adding descriptive clarity to CVR's other interpersonal resources.

The findings are framed as the 360° image's impact on potential interpersonal meanings, which have an ethical dimension when the texts are representing real social actors.

1.3.6 Chapter 7: Information Organisation

This chapter describes the way CVR structures its information visually, multimodally, and temporally across sequences of shots. Like the previous chapters, this chapter is broken into two main sections. The first section addresses the 360° image for its information structures. The second section discusses how these structures become incorporated into the system of conjunction and cohesion, where multimodal meanings proliferate logically across larger tracts of text. The second portion of the chapter differs in that it presents two macro-analyses of the texts from the perspective of cohesion.

1.3.7 Chapter 8: CVR As Situated Discourse

This chapter discusses the texts at a higher contextual level than the previous chapters. I present the findings from a register analysis of both texts, framing them as instances of nonfiction discourse. Where previous chapters describe CVR at low levels of meaning, this chapter takes a holistic view, presenting CVR as the confluence of technological and semiotic properties. The texts

are described for the different approaches taken in their management of the immersed spectator.

The chapter also functions as a culmination of the findings across chapters five to eight and is where I present my main thesis. My findings are contextualised against the deterministic claims to reality, made on CVRs behalf. In doing so I will draw on the concepts of dramaturgy and performativity outlined in the literature review chapter.

1.3.8   Chapter 9: Conclusion

This final chapter addresses the research across the three strands of the research question, across low-level and higher-level findings. As a semiotic technology, CVR is discussed as low-level discourse that affords a degree of reflexivity and autonomy, but which is always constrained at higher levels by the authorial construction of the text. Nonfiction CVR production is then addressed along these lines of autonomy, control as well as the semiotic independence of filmed subjects in CVR. Discussion here also includes practical considerations for producing nonfiction CVR texts. Lastly, the chapter addresses the study from the perspective of process; as the theoretical and practical considerations required to study discourse in nonfiction CVR texts.

2. Review of the Literature

The chapter is divided into three sections. The first section examines literature useful in conceptualising CVR as a novel emerging technology. I present a number of genealogical media theories and present Bolter and Grusin's *remediation theory* (1999) for its usefulness as a frame for discussing CVR's technological hybridity. The second section 'decomposes' CVR, using remediation as a loose frame for considering CVR as the confluence of other media forms. The third section addresses literature on nonfiction, beginning with a broad exposition of nonfiction theory, highlighting its complex semiotics, and ending with nonfiction CVR-specific literature. I compare critiques of nonfiction CVR with those from earlier movements. In this section I also present a dramaturgical frame used to describe nonfiction, which is used in broad discussions on CVR, in later chapters.

2.1 CVR: An Emergent Hybrid technology

The emergent nature of CVR can be framed within genealogical meta-theories. Two complementary theories, the 'double-births' model developed by André Gaudreault and Philippe Marion (2002) and remediation theory (Bolter, J. D. Grusin, 2000), can provide high-level context for the nature of CVR's emergence as well as characterising its current phase of development, and the technological claims made on its behalf.

2.1.1 The 'Double Births' Genealogical Theory

Gaudreault and Marion model cinema's evolution from its technological inception to its maturation as an autonomous form replete with its own institutions. According to their model, a media form emerges in two phases: the *first birth* and the *second birth*. They characterise early cinema's first birth as a technological phase, where novelty drove interest in the form, until it reached its second birth when it became a stable form with its own institutions, practices, styles, and aesthetics. The early cinematograph (the moving film image) was a technological rupture that subsumed, yet substantially altered, pre-existing technological forms (ibid) such as the chromatograph, bringing about a new representational paradigm. Where the chromatograph *illustrated* movement through sequences of static frames, the cinematograph *showed* movement, presenting it on screen to the extent of defying its own physical makeup (of fixed image frames on a plastic strip). It is this novelty that sustained cinema throughout its early developmental phases. In this regard, CVR can be considered as in its first birth, as a technological rupture that is novelty driven. CVR's subsuming of filmic image capture into VR

display has driven the interest in the form since its inception, as Hollywood filmmakers devoted their resources to understanding CVRs potential (e.g., Katherine Bigelow (2017); Spike Jonze (Vice News, 2015)) as well as scholars (Uricchio et al., 2016).

An aspect of Gaudreault and Philippe Marion's model that is familiar from CVR's perspective is that the first birth, being largely technological, relies on extant media systems for its texts as it 'gets to grips with pre-established codes' (Gaudreault and Marion, 2002, p.12). Indeed, the earliest phases of cinema were typified by its subordination to other performance genres, as a 'cinema of attraction' (Gunning, 1986) subsumed into and incorporating Vaudevillian entertainment genres (Keating, 2014). Like early cinema, CVR is frequently used to allow a novel technological access to other kinds of genres. These include circus performances (Cirque du Soleil: Dreams of 'O' Lajeunesse, 2017) and dance performance (Together: Malick, 2018) as well as a raft of amateur producers using CVR to present novel experiences in non-performance contexts (e.g., skydiving (360° VR Skydive for Next Level Indoor Skydiving, 2018)).

The model, while somewhat predictive, is simplistic and omits the complex socio-economic aspects that drive (and obstruct) the development of media technologies. Other criticisms, such as by Rouse (2016) consider it as 'media-centric' and teleological where essentially the winner writes history, relegating other forms to 'primitive or naive version[s] of the cinema we know today' (ibid., p. 97). This notwithstanding, the model provides high-level historical context and a loose frame in which to track media's evolution from technological fancy to an established form with its own identity.

2.1.2   Remediation Theory

Bolter and Grusin's Remediation Theory (2000) provides a more nuanced genealogical account of media's evolution. The theory is similar to Gaudreault and Marion's model where new media are described as subsuming old, restating McLuhan's maxim "the content of a medium is always another medium" (2000, p.65). It differs in scope as an explanatory metatheory for both the emergence and survival of all media forms. Remediation is described as a process of absorption and adaptation, between technological forms where media forms mimic and 'refashion' each other to establish themselves as viable in the contemporary media space. These two dimensions of the theory (emergence and survival) are co-extensive in the larger remediative frame but as I am concerned with the emergent nature of CVR, I will foreground the aspects of the theory that allow for discussion of CVR as the remediated parts it inherits and refashions. This will involve obvious remediations between VR and film, as well as other non-technological forms such as immersive theatre.

9

Remediation also has its critics, which I will discuss at the end of this section, but it avoids claims of teleology and media-centricity in its basic tenets. Rather than positing a linear movement towards an ideal state, media forms are considered to be in perpetual flux, in *dialectic* with each other as they absorb each other's ontologies. The flux is ordered however by two overriding 'logics' that organise the evolution of new media forms: as *transparent immediacy* and *hypermediacy*. Transparent immediacy operates on the erasure of the media form, endeavouring to negate the experience of mediation. When erasure is successful, the display mechanism becomes 'transparent', providing a more direct contact with that which is mediated. Dobson terms this as a 'window through' media (2009). Film, gaming, and VR exemplify this logic as they present seamless perceptual spaces as extensions of our own immediate embodied space. As an embodying logic, texts that display transparent immediacy are inherently immersive (Bolter, J. D. Grusin, 2000). For the authors, the embodied self is *mobile* and *present* in the mediated world, leading to heightened emotional states such as *empathy* (ibid). They use the fictional sci-fi film *Strange Days* (Katherine Bigelow 1995) as a hyperbolic 'end game' for this logic where media content is captured directly via human experience, stored and replayed on a device called 'the wire' which is plugged directly into the spectator's mind. The experience is in this sense unmediated as it is transferred directly from human to human. Through *Strange Days*, the authors articulate the underlying desires of CVR production, namely, to give the spectator the unmediated experience of reality.

The second logic of *Hypermediacy* assumes for the most part that instead of a unified transparent space, we are confronted starkly with the opacity of mediation. Fragmented, media-rich web interfaces and contemporary news TV platforms are used as evidence of this kind of remediation where the underlying experience becomes one of multiplicity with multiple modes of access to information (the 'window through' metaphor becomes a 'window-at' (Dobson, 2009, p.2)). Bolter and Grusin maintain that hypermediacy of this kind leads to a "feeling of fullness, a satiety of experience, which can be taken as reality" (2000, 53).

The logics are considered as phenomenological states that can at times co-exist in single texts. The authors cite the apocryphal stories of audience response to the Lumiere's *Arrivée d'Un Train* (Arrivée d'un train (à la Ciotat), 1895) as having components of both logics, producing a form of dissonance; of seeing something 'real', yet undeniably mediated (Bolter and Grusin, 1999, p.155). Dissonance in contemporary film-viewing is produced "through the use of stylistic techniques that violate Hollywood conventions": essentially, where Hollywood's normative and 'natural' devices (read: classical narrative continuity) for transparency are transgressed, either

unintentionally or for stylistic reasons, the viewer is drawn out of the experience of immediacy. This is apposite to studying CVR which, through its immersive tropes, makes the spectator aware of their own presence (Kool, 2016b), suggesting a dissonance in their overall experience. For example, CVR texts that give the spectator a sense of being (or playing) a character, replicate the filmic device of 'breaking the fourth wall', which is considered an unnatural look (Gunning in Brown, 2013), that is as such hypermediated. I will return to this aspect of CVR through a more detailed exposition of film and VR as immersion-based technologies by exploring the psychological phenomenon of presence in section 2.2.1, *CVR as Hybrid of Film and VR.*

Criticisms of remediation centre on the deterministic nature of its logics, a point the authors are keen to debate. They accept that their adherence to McLuhan runs counter to the prevailing ideologies of the time, as espoused by Raymond Williams (cf. Lister, 2009), but they perceive remediation as a means of addressing what they see as an imbalance towards Williams' accepted dogma. They state that "We need not be afraid of McLuhan's 'formalisms,' as long as we remember that technical forms are only one aspect of technologies that are simultaneously social and economic" (1999, p.77). As such, they tread a centre-ground, accounting for the economic impetus for media development.

2.1.3 Concluding Remarks

Remediation is sufficiently high-level to be used in this study as a descriptive frame for CVR's hybridity, by discussing CVR as a dialectic between multiple forms. The logics also frame some of the analytical theory used in this study. For example, transparent immediacy, as illustrated in *Strange Days* touches on some ethical and epistemological aspects of nonfiction CVR, of immediate contact with social reality and social actors. Also, the kind of sensory experience the transparent logic aspires to is conceptually aligned with the semiotic concept of 'bandwidth', as the technological mode of sensory contact that media forms afford. The criticisms of in-built technological determinism are reasonable but as I am taking a social semiotic approach, the study is inherently focused on the social uses of CVR. Using remediation's logics and the tenets of remediative dialectics, I will examine CVR as the sum of its parts by decomposing it into film, VR, photography, and theatre.

2.2   Decomposing CVR

2.2.1   CVR as Hybrid of Film and VR

VR and film share a common transparent lineage from CVR through VR, film, photography, camera obscura, panoramas, and perspectival painting. They each reformulate the same desire to *transport* the viewer through realism and self-erasure. Their similarities notwithstanding, film and VR have tended to inhabit distinct theoretical domains. As an older medium, film has a long theoretical legacy spanning multiple eras and 'isms' and has been conceptualised in diverse fashion, from positivism (cf. Vertov published in Michelson and O'Brien 1984) post-structuralism (cf. Renov, 1993), psychoanalytical (cf. Cowie, 2011) to phenomenological frames (cf. Sobchack, 1990; 2004). VR on the other hand, being predominantly interactive and computer graphics- based has largely been produced and studied by computer scientists and engineers. Classic VR became concerned largely with *human computer interaction* (HCI), with different scholarly traditions (Jerald, 2015). Although the idea of virtual reality and *virtuality* more broadly has captured the imagination of scholars and popular culture (cf. Irwin, 2002) the canon of research into VR production and reception has taken a predominantly behaviourist approach in discerning 'successful' psychological VR experiences (cf. Laurie M. Wilcox, Robert S. Allison, Samuel Elfassy, 2006; Bouchard et al., 2008; Coxon, Kelly and Page, 2016).

The film image was theorized in the early and mid-Twentieth Century theorists for its formal attributes and experiential qualities. Metz (1974) differentiated the film image from other modes of representation in its ability to generate, through projected, moving, and indexical means the *impression of reality* which formed the basis of perceptual and affective responses. In this light, film functions on a *credible* level; the impression of reality is the basis on which film opens the viewer to further narrative, rhetorical and ideological persuasion. Part of this impression, for Metz comes from the photographic realism achieved by the camera, and the illusion of movement achieved by projection. To illustrate his point, he elaborates Barthes' musings on the printed photograph which is always 'here' in our hand as a photographic object, while the historical trace of what it represents is always gone ("the illogical conjunction of here and then" (ibid. p.2)). Even though filmic frames are acquired in virtually the same manner as printed photographs, their sequential display, coupled with our persistence of vision generates movement that is perceptually happening 'now' for the viewer. In other words, to *look at* a photograph is to position oneself once removed, temporally, from its representation, whereas to *watch* a film is to locate oneself co-extensively with cinematic time. For van den Berck this amounts to life-like quality where "…motion imparts *corporeality* to objects and gives them

autonomy their still representations could not have" (Metz, 1974 p.2, my emphasis).

Regarding *transportation*, in film it is the viewer's immersion into the cinematic space, constructed through continuity editing; in VR it is the continuous psychological experience of *presence.* Presence is described by Slater (2003) as a subjective phenomenon that is contingent on technological immersion, enabled by sophisticated hardware and software as well as content verisimilitude (Ibid.). The sense of presence is predicated largely on erasure, as "our awareness of the medium disappears and we are pushed through the medium to sensations that approach direct experience" (Biocca, 2002, p.102). Presence has also been termed as 'place illusion', as "the qualia of having a sensation of being in a real place" (Slater, 2009, p.3549). A separate yet related phenomenon is 'plausibility illusion', "the illusion that the scenario being depicted is actually occurring" (Ibid.). The latter is a valuable distinction to make in that it shifts the emphasis on to what *happens* to the spectator once they have been successfully transported to the virtual world, a concern at the root of this study.

Computer generated VR and CVR share a common display format that erases itself allowing the spectator to rotate and survey an immersive image. Computer generated VR, however, has a claim to a kind of phenomenal realism that CVR does not, owing to the way its content is generated: where CVR *represents* reality, VR *simulates* reality. Lister (2009), writing before CVR's inception describes this dichotomy as occurring between film and VR, where filmic representation *mimics* reality and VR *models* a code-based reality facilitating different kinds of behaviours on the part of the user (Ibid). As it is built on code, computer generated VR facilitates 'six degrees of freedom' enabling the spectator to rotate their view around horizontal and vertical axes, but also to pivot around the 'z' axis, an action CVR spectators cannot carry out. Critically, users of VR can move around the scene in all directions ('translational movement' (Barnard, 2019)). VR therefore allows the user to feel that they have access to limitless horizons of spatial experience. These will naturally be constrained within system limitations, but the experience will be designed to obscure this fact. This translational freedom to move highlights a remediative dialectic between VR and gaming where both, in the Heideggerian sense, provide the user a particular 'ready-to-hand' kind of engagement (Fortner and Fackler, 2014).

2.2.2   The Camera as Proxy

Where VR allows a kind of 'freedom to roam' within its models and simulations, any roaming done in CVR, like film, is determined by the position and movement of the camera. A point of view is 'given' to the CVR spectator much like in framed film but with added frameless-ness and rotational aspects. Perspective is therefore core to understanding how CVR functions to make meanings for the embodied spectator and can be discussed theoretically in the context of filmic

perspective.

In any lens-based representation, the viewer sees only what the camera saw, and through a process of identification, these two views merge as filmic perspective. Ponty (in Shaw, 2008) expressed this notion of camera-as-proxy as the phenomenology the film experience. For Ponty, the camera-body-eye relation positions the viewer temporarily in a 'virtual body' via the perspective of the camera. Essentially, our body is a primary component of experience, of immediate bodily perception in the phenomenal field that is pregnant with meaning, yet to be granted *significance* through subjective perspective ("condemned to meaning" (p.61)). The camera, as an embodied and seeing thing, *extends* our seeing - as *projected* for us to see. Ponty states: "Film differentiates itself by being an act of seeing that makes itself seen, an act of hearing that makes itself heard, an act of physical movement that makes itself reflexively felt and understood" (Ibid, p.62). Benjamin (in Cohen, 2001) described this filmic identification as happening on two levels; first with the camera, where our visual perspective is subordinated to that of the camera's; secondly, identification to the character to whom our visual attention has been directed (2001).

Cinematic meanings rely on the camera being a proxy for the spectator's vision, as it provides a defined point of view (POV). The fusion of camera and POV means that the *positioning*, *movement*, and *angling* of the camera's lens, all will engender an equivalence of *place*, *motion,* and *orientation* in the viewer. Here, the experiential aspect of film becomes meaningful through cinematography and editing. POV is the basis for compositional decisions typically expressed as *framing*. Filmmakers use framing to make their scenes comprehensible, where the spectator knows 'where they are' in the unfolding action, and has a view on the filmed subjects, affording emotional potential. Generic shot types evolved over cinema's history to provide distinct embodied POVs (for a compendium of shot types see Brown and Safari, 2016). A 'long shot' for example frames the subject at a distance (either literally or through lens manipulation ('zooming')), showing context, locating both the subject, and the spectator's view. 'Close-up' shots use framing to bring the viewer and viewed subject into close contact, allowing detail on actions, reactions, and emotions. High and low camera angles, when paired with close-ups engender a relative power between viewer and subject. The camera-as-proxy positions the spectator's view 'looking down' or 'up to' the subject, giving the perception a degree of domination or subjugation. In this regard, POV is often conceptualised as a dichotomy between 'objective' and 'subjective shots' (Lancaster, 2019; Brine, 2020). A distant view, for example, establishes a more neutral and objective view on events compared to the more subjective close-up that positions the viewer and subject notionally in a conversational, or more intimate

14

arrangement [1]. Where all the shot examples provided here afford a notional POV of sorts, where the viewer is an invisible presence in the scene, 'first-person' POVs are held as the prototypical subjective shot, where the camera takes the place of an actual character in the shot. These POVs are stark and hypermediated as they draw attention to themselves.

Branigan (1984) provides an elegant conception of filmic POV that is perhaps more useful in discussing the kinds of perspectives afforded in CVR. Branigan eschews 'objectivity', instead conceptualising film as a series of *nested subjectivities*. At the highest level it is the *viewers'* POV, through which we access the *filmmaker's* POV. In turn, the filmmaker may present us (via the camera) with the *character's* POV. The essential difference for Branigan is that where shots are considered 'objective' they are nonetheless presenting *someone's* POV, which in the absence of a recognised character-narrator, is necessarily the filmmakers. Conversely, when the filmmaker decides to present the narrative through the POV of a character, the narrative unfolds from 'his or her point in space' (Branigan, 1984, p.2). This conceptualising of filmic subjectivity is particularly useful in the nonfiction context where perspective often takes an ethical and epistemological charge.

Branigan's theories were developed to describe fictional film perspectives, but they are apposite as a frame for this study in making sense of CVR's reformulating of POV and perspectives. In a broad sense, concepts of objectivity, neutrality, and subjective POVs require reconceptualising in CVR. CVR usually presents perspectival choices of either first-person or second-person *bystander* POV, elsewhere described as 'actor' and 'observer' (Larsson, Västfjäll and Kleiner, 2001; Anneliene et al., 2015) which all fall within the subjective range of conventional filmic POV. Some have gone as far as to suggest that there are only first-person experiences available in CVR (Laurel in McRoberts, 2017; Raphaël in Tricart, 2017). What is notable considering Branigans concepts is the apparent technical fusing in CVR of his top-level *viewer's* POV with the other camera views provided by the films.

CVR also must negotiate the *temporal* aspects of film and VR in its remediating dialect, where film requires continuity across its edited shot sequences to maintain transparency and the illusion of transportation. Bolter and Grusin reference the '180° rule' as a Hollywood standard in continuity editing filmmaking that if broken would alternate the experience between transparent immediacy and hypermediacy (Bolter and Grusin, 1999, p.156). This rule is part of a complex and sophisticated set of 'invisible editing' practices (Passmore et al., 2017) involving

---

[1] I am presenting these as 'the basics' to illustrate fundamental aspects of framing. There will inevitably be exceptions to these 'rules' that I will not get into here.

shot framing, screen direction, as well as spatialised sound. The rule dictates that the camera must stay on one side of the action, ensuring coherent *screen direction*, where objects appearing screen-left and right do not 'flip' back and forth (making the viewer feel like they are the ones 'flipping'). The result of breaking this rule is a 'jump cut' that draws the viewer out of the stream of continuity and makes them aware of the process of mediation. 'Screen-left' and 'screen-right' are of course less stable in CVR where the spectator rotates their view to accommodate objects of interest. Where continuity is created across tightly edited shot sequences, the longer shots in CVR contain their own indexing of continuous activity in what are usually longer shots.[2] Where the capacity for jump cuts is reduced, the problem of continuity becomes related to maintaining the spectator's attention in the specific flow of activity which is pertinent to the story. Continuity in CVR therefore becomes the perennial concern of successfully *directing the spectator's attention*, where the goal of production theory is in finding ways to guide the spectator through narratives in covert ways (cf. Brown et al., 2016b; Dooley, 2017b; Elmezeny, Edenhofer and Wimmer, 2018).

The broader question that emerges through studies such as by Passmore et. al (2017) is whether a linear through-line is in fact the most suitable means of presenting stories in CVR, and whether it is compatible with the ways people attempt to make sense of immersive texts. In their study they suggest that the natural tendency is to explore and make ad hoc judgements concerning when to focus on entities and when instead to look around (Passmore et al., 2017, p.9). In a wider sense, the imposition of a linear narrative in a 360° experience harks back to Gaudreault and Marion's 'first-birth', as CVR is subjugating itself to linear framed paradigms. From a remediation point of view, this reflects the concept of 'weak remediation' put forward by Bolter and Grusin, where CVR is transplanting of one kind of textual engagement into another. These linear/nonlinear issues are teased out in detail in this study from a semantic point of view, suggesting that CVR lends itself to active ad hoc constructions of logical meanings.

2.2.3   CVR and Theatre

To this point, CVR has been described largely along filmic lines together with the rotational contribution of the headset. Theatre can also offer insights into CVR and the way it differs from framed film. First there is the overarching hypermediated nature of theatre, where the actors

---

[2] This is contested by some researchers: Kjaer et al (2017) suggest that shot frequency is not detrimental to CVR viewing which is countered by claims that it takes the spectator out of the experience (Godde in Dooley, 2020, p.84). Both texts used in this study use long shot durations.

on stage are 'too real' and a barrier to the kinds of identification found in film (Metz, 1974). There are resonances in this regard with CVR, which also presents hyper-real situations with which the spectator is reflexively located. Secondly, regarding immersive theatre, there is a shared impetus to position the spectator inside unfolding action. Nonfiction CVR's claims to realism are predicated on direct contact, which is mirrored in several socially oriented immersive theatre experiences.

Haedicke (2002) describes the immersive theatre piece *Un Voyage Pas Comme Les Autres Sur Les Chemins de l'Exil* (UNHCR, 1999) as a 'mise-en-situation' which embodies the spectator as a notional but specific, named refugee in their arduous stages of their refugee journey. The press release for *Un Voyage* reads: "By adopting the identity of a refugee who flees his/her country, which is subject to war, persecutions, or dictatorship, the visitor discovers all the stages of a request for asylum" (Ibid, p.102). 'Adopting the identity' reflects the 'actor' POV considered by some as the preferred view in CVR (Anneliene et al., 2015). Other immersive theatre productions replicate this embodied approach such as *Pinjare* (*Cages)* (Dinesh, 2016), which was produced for a male Kashmiri audience who adopt a gendered perspective in order to 'see' first-hand the prevalent hierarchies that exist in their society. Here, the spectator is asked to take on the perspective of a Kashmiri woman to "embody the gender-based differences in how the region's conflicts are experienced" (ibid, page 2). The effect of embodiment intended here, is to foster a sense of identification and empathy with the person whose experience they are inhabiting. Cohen (2001) describes identification as "an imaginative experience in which a person surrenders consciousness of his or her identity and experiences the world through someone else's point of view" (ibid p.248). In general psychoanalytical terms, identification is the process of a subject merging psychically with an object (the parent, social peer, aggressor for example) to assimilate some trait beneficial for personal development, social identification and ultimately, safety. In essence, identification requires that "we forget ourselves and become the other" (ibid p.247), a phenomenon that carries across into media contexts, such as through films' absorptive realism (Ibid.). The expectation would be that by taking the place of the subjects (refugees in *Un Voyage Pas Comme Les Autres Sur Les Chemins de l'Exil*, and with the Kashmiri women in *Pinjare*) as 'identification targets' (Ibid, p.251), psychic merging would become context specific, as the immersive theatre participant fleetingly negotiates the subjects' trials and tribulations.

A similar mindset existed in CVR's earliest phase: that immediacy automatically engenders empathy. As a result, a high proportion of productions aim to transport the spectator to crisis

zones and sites of human despair.[3] Films such as *Clouds Over Sidra* (Milk and Arora, 2015a) *Waves of Grace* (Milk and Arora, 2015b) and *My Mother's Wing* (Arora and Palitz, 2016) were produced in partnership with the *United Nations Virtual Reality project*, explicitly with the stated intention of 'pushing the bounds of empathy' (United Nations, 2018) to which end, Chris Milk gave CVR the now infamous epithet of 'empathy machine' (Milk, 2016). The fundamental difference between the claims made by these films and those made in the immersive theatre examples above is in the level of identification, where CVR more frequently invites the spectator into an 'immersive witnessing' paradigm (Gregory, 2016; Nash, 2017), rather than as an identified participant.

On an epistemological level, immersive theatre has been differentiated from other texts for its multisensory nature (Machon, 2009; White, 2012). For White, theatrical immersion distinguishes itself by rebalancing perception from the object back to the senses themselves and "address[ing] multiple senses simultaneously, rather than addressing itself primarily to conscious thought" (2009 p.228). For Machon, this synesthetic engagement with the text "…means to perceive the details corporeally" (Machon, 2009, p.17) and collapses 'sense' as meaning making and 'sense' as sensation and emotion into one 'site of response' (ibid.). This stands to reason as immersive theatre activates all senses simultaneously. It also points to the ways film, VR and CVR present sensory experiences that differ from immersive theatre quantitatively, and from each other qualitatively. Film's corporeality, for example has been described by Sobchack (2004) for its inherent sensuality, and more recently the rise of *Autonomous Sensory Meridian Respon*se (ASMR) (Poerio et al., 2018) in popular consciousness points to the tangible bodily effects of audio-visual media, beyond the aural and visual senses. VR studies have shown that the neural effect of erasure (Riva, Wiederhold and Mantovani, 2019) where normal sensory inputs are supplanted by VR representations yet the 'body matrix' responds as-if VR's synthetic inputs were real. Sutherland (2012b) described an 'exteroceptive' sensitivity to the occluded elements of VR scenes. When CVR's verisimilitude is factored in, these effects are in response to actual scenes and in this regard mirror immersive theatre albeit with far more tenuous sensory connections to the 360° image and where the spectator is 'bodiless' and akin to a floating consciousness.

2.2.4   Decomposing CVR: Concluding Remarks

A picture of CVR emerges through the lens of remediation, where its similarities and differences

---

[3] The VR Mediography comprises a total of 603 titles, at the time of writing 81 titles are themed. 'war and conflict' and 43 as 'migration and displacement', amounting to roughly one quarter of the total productions

with its remediated forms become its formal characteristics. Filmic realism becomes remediated yet complicated in dialect with CVR which has no frame and picture plane. CVR's indexical moving representations are no less real in terms of their 'corporeality', but the experience of seeing them in CVR is phenomenologically different, where the spectator's own reflexive corporeality is included in the experience. Conventional filmic transportation into the conventional (framed) cinematic space is predicated on the viewer momentarily leaving their physical space (Wallon in Metz, 1974). In CVR, the mechanics of viewing the films in the head-mounted display (HMD) results in a collapsing of the picture plane, where image-space and reflexive-place are fused. The ramifications for this are found in the kinds of resources that are absorbed and transformed through remediation. For example, conventional framed lenses afford the filmmaker the ability to frame things *as if close*, through the optical mechanics of the lens ('zooming in'), with an intimacy that is not taken literally by the viewer, owing to the distancing effect of the framed picture plane. On the other hand, in CVR, distance is indexed to the distance between the camera-spectator and subject (Dooley, 2020). Closeness is experienced in a more literal embodied fashion, and the traditional close-up shot becomes problematic in CVR, and even unsettling (Unseld, S in Cone, 2015; Hamlin in Tricart, 2017, p.107). Additionally, vertical angles (as relative power) are predicated on the traditional camera's single unidirectional lens that can be directed upward or downward. The CVR camera is omnidirectional and can only be positioned above or below the subject, facilitating a *likely* relationship of power but not necessarily *instantiating* one.

Regarding VR and theatre, other significant junctures present. For VR it is the loss of the additional dimensions of freedom, where the ability to 'roam' becomes constructed to a fixed position in pace. The trade off, can perhaps be the nature of the low-level semiotics of CVR, where an *indexical* connection with reality is available, in contrast with *iconic* computer-generated forms. Regarding theatre, similarities are found in CVR's multi-sensory reception and necessary mode of production. Like immersive theatre, CVR activates an embodied multisensory response to the environment, but only in a synthetic technologically mediated fashion. As a mode of production, similarities are found in the level of staging required. Where the omnidirectional CVR camera captures in all directions, there is no 'behind the camera' for the filmmaker to manage and respond to the scene in real time, nor for the represented subjects to interact with the filmmaker. As such, when the subject is left alone with the CVR camera, a sort of theatrical performance to the CVR camera is required (Kool, 2016b). Essentially, the conventional camera, being more manipulable and mobile can respond to the scene, whereas CVR production requires the subject to orient their activities to the CVR camera, requiring a fundamental shift in practices.

2.3 Nonfiction

2.3.1 Nonfiction Genre

'Nonfiction' is a somewhat contested label, comprising a complex taxonomy of sub-genres. This is made more complex by the proliferation of reality television genres (Corner's 'post-documentary era' (2000, p.687)), and the diffuse array of nonfiction new media texts that has emerged in the last two decades (for a comprehensive list, see MIT, 2019). This subsection first establishes a working definition of genre for the purposes of this study, which is then used in this chapter as a frame to review literature relating to nonfiction genres.

Martin and Rose provide a functional definition of genre as 'staged, goal-oriented social processes' (Martin and Rose, 2008, p.6). 'Staged' in this context refers to the way a text is assembled: a narrative genre will involve an *orientation*, *conflict,* and *resolution*, (Labov, 2013) with optional stages such as a *coda;* a *report* genre will involve the decomposition of a thesis or topic into a series of evidentiary stages. Where some text types realise one of these genres exclusively, others, including documentary texts, will mix these generic patterns. In this section I am interested primarily in the nature of the texts' '*goal oriented social processes'*, which in the context of nonfiction, have driven much of the last century's theoretical debates. Broadly speaking, the goal-oriented processes that nonfiction texts perform share the need to present something of the real world to an audience; informing and engaging in the manner summed up in Griersons maxim '*the creative treatment of actuality*' (in Winston, 1995). In documentary production this manifests as the deployment of two modes of thought, described by Bruner as the *logico-scientific/paradigmatic* (rhetorical) and the *narrative* (Bruner, 2009). For Bruner, the logico-scientific mode is the scientific, logical, and empirical mode of thinking that takes as its broadest aim the logical reduction of the *particulars* of what is communicated, to an underlying and verifiable *truth*. The *narrative mode* on the other hand is inherently experiential; stories require us to imagine things from a human perspective, "in human or human-like intention and action and the vicissitudes and consequences that mark their course" (ibid. p.13). The 'social' aspect of genre reflects the social and ideological norms that determine and uphold genre. To claim 'documentary' status for example implies an authentic engagement between the viewer and the world represented. Where a text is indexed as *nonfiction* and subsequently found to be misleading or unbalanced in its representations, it subverts its genre, and the film may be rendered redundant. *Searching For Sugar Man* (Searching for Sugar Man, 2012) is such a film which omitted critical information about its subject, presenting a quasi-fictional account of the

subject, an aspect of the film that has become a part of the film's legacy (Titlestad, 2013).

2.3.2   Fiction and Nonfiction: A Continuum of Practices

The line between nonfiction and fiction is not always clear, and theories suggest they inhabit a continuum of shared practices, making differentiation difficult. Nichols (2001) probes the distinction thus: "documentaries address the world in which we live rather than a world imagined by the filmmaker" (Nichols, 2001, p. xi). With this as his premise, he elaborates the ways in which 'worlds imagined' overlap with the historic world, as with biopics where the fictional world represented on screen has its provenance in actual events. Stella Bruzzi (2000) uses *Dockers* (Dockers, 1999) as an example of a fictional film that is based on a historical event (a docker's strike), ostensibly similar to a straightforward biopic but produced in such a way that reality reaches further into the actual production of the film. Dockers' fuses reality and fiction in its screenplay which was developed in collaboration with the actual strikers, who also performed on-screen roles in the film, allowing a curious mix of biography and fiction.

Nonfiction has been described more systematically for the way it shares a common set of resources with fictional films. Nichols describes how documentary films address the multifaceted and complex historic world through linear storytelling, using conventions native to fiction such as *staging, re-enactments* and *performance* (Nichols, 2001, p.38). Renov describes the semiotics of fiction and nonfiction as 'enmeshed in one another' with textual codes sharing more similarities than differences. "Poetic language […] narration […] musical accompaniment to heighten emotional impact" (Renov, 1993, p.2) are common to both fiction and nonfiction with the distinguishing feature being the 'historical status of the referent' (ibid.). Indeed, many of the classic documentaries of the early twentieth century employed the resources of fiction film to represent their subject matter, adopting the studio practices of the time. *Night Mail* (Night Mail, 1936) for example had many of its interior shots filmed on a sound stage (Nichols, 2010) producing a dramatic yet unnaturalistic style typical of early twentieth century cinema. Early documentarians such as Joris Ivens (Power and the Land, 1940; La Seine a rencontré Paris, 1957) saw to it that they extract 'natural performances' from their filmed reality subjects, using extreme contrivances to construct a narrative arc replete with conflict and resolution, not unlike their Hollywood contemporaries (Waugh, 2011). In *Power and the Land* (Ivens, 1940), the main plot involves a farming family becoming part of a Rural Electrification Scheme. For practical reasons Ivens used a farm that was already electrified prior to production (Smith, 2019), requiring him to conceal the electrical items at the beginning of the film so as to reveal them towards the end. The cinematic world in *Power* has very weak claims to authenticity by

contemporary standards, and its mode of production was found in other contemporaneous productions such as Flaherty's engineering of his subjects' performances (e.g. Nanook of the North, 1922; Moana, 1926; Man of Aran, 1934).

2.3.3 The Evolution of Nonfiction Forms

Nonfiction genres have evolved since Ivens and O'Flaherty's productions, reflecting ideological and societal shifts, as well as advances in production technology. Nichols tracks the broad milestones in documentary practice and reception through his documentary 'modes' (Nichols, 2010). He points to six modes:[4] *Expository, poetic, observational, reflexive, participatory* and *performative*. As a metatheory, his modes are open to similar criticisms as those levelled against Gaudreault and Marion - as a teleological reading of documentary's evolution (Bruzzi, 2000) but while they are admittedly limited, they remain a useful point of historical orientation for many contemporary studies in nonfiction and documentary (cf. Castells, 2011; Gaudenzi, 2013). Nichols' modes are predicated on the notion of *voice* as heard in the discourse between author, viewer, and filmed subject. Voice can be considered both literally and figuratively in terms of author and subject, where they may each be given agency in terms of their voice is being heard (as voice-over or social actor) and their *mode of address* (Nichols, 1983). For Nichols, voice becomes figurative where the heard voices in the film are superseded by the authorial voice of the filmmaker, manifesting in the 'organizational strategy', or how they speak *through* the film (Nichols 1983a, p. 20).

The seminal *expository* and *poetic* modes stem from the early Griersonian era, typified as the 'voice of God' style of authoritative and didactic voice-over narration. They differ in their tone and affective qualities, from factual to poetic with both employing scripted rhetoric intended to bring the viewer into line with some sort of message. The World War 2 film series *The March of Time* (The March of Time, 1935-1951) is held up as an unambiguous illustration of this didactic voice (Nichols, 1983) giving the author unequivocal standing, constructing a passive viewer. Ivens' *The Power and the Land* (1940) is no less didactic: the mode of address heard is 'folksy' playing on romantic ideas of rural America, yet to modern ears its folksiness becomes its subterfuge. The didactic voice takes on a sinister ethical dimension in scenes involving interaction between characters, who's dialogue is seen but not heard as the voice-over is transposed over their unheard conversation. There is an inescapable sense in moments such as this of the filmmaker putting words in the mouths of those on screen, a practice that would become anathema to

---

[4] For brevity I will take advantage of the overlaps within certain modes: performativity and participatory modes for example share much conceptual ground relevant to discussions.

documenting social situations in the observational movements that followed.

The *observational* modes of the 1960s and 1970s evolved as a confluence of technological advancement, within postmodern ideas of authorship that rejected didacticism as a voice of deeper modernist metanarratives. Observational filmmakers include Frederick Wiseman (Hospital, 1970), the Maysles brothers (Salesman, 1969; Grey Gardens, 1975) and Robert Drew (Primary, 1960) who adopted a transparent and immediate style enabled by lighter portable cameras and synchronized sound recording (Rothman, 2004, p.281). Filmmaker and crew could move freely and invisibly in the subject's space and voice-over could be replaced with location sound, ostensibly allowing the subject to reveal their own truths, in their own voice. The observational styles have become a locus for debates around nonfiction regarding notions of erasure and transparency which reverberate in CVR's early hype cycle (Milk in Dredge, 2015). As such, I will return to this later in this chapter (section 2.3.5 *Camera Technology and Discourse*), making more detailed comparisons between CVR and the observational style.

Subsequent documentary modes that emerged roughly contemporaneously to Nichols' time of writing were typified as *reflexive, participatory* and *performative*. For Nichols the performative mode raised questions of subjectivity, of 'what knowledge amounts to' and 'what counts as understanding and comprehension' (Nichols, 2010, p.201). Interpretive, subjective representation was forefronted, over earlier Cartesian epistemologies rooted in positivism, where knowledge is 'abstract' and 'disembodied' and can be 'exchanged' in its transmission from author to viewer (ibid.).

2.3.4   Nonfiction and Performance

Performance has become a prominent concept in nonfiction discourse, grounded in J.L. Austin's theories, specifically his *performative speech act* (Austin, 1975). Essentially, a speech act is an utterance that is itself a self-contained action that constitutes its own reality, such as the "I do" part of a wedding ceremony (ibid. pg. 5). Austin's theory is invoked by Nichols and Bruzzi to characterise nonfiction performance (Bruzzi, 2000, p.186; Nichols, 2010, p.203). Nichols stops short of describing performativity in nonfiction as something that is *generative* of a reality, a point that Bruzzi is keen to address (2000, p.186). For Bruzzi, performativity is not an *aspect,* but a *necessary condition* of documenting. Regardless of the specific events at hand in the documented situation, the *reality of documenting* can only unfold as the sum of its performed 'utterances'.

The performative aspect further problematises definitions of nonfiction that take the notion of 'reality footage' as the basic differentiating premise. For Renov, nonfiction could look like fiction

but could always be distinguished by the provenance of its footage, or '…the *'historical status of the referent'* (1993, p.2). When performance is foregrounded, the 'codes and conventions' of fiction reach further back into the production of the footage, where filmmakers consciously become actor and character, as much as author. A documentarian of interest to both Nichols and Bruzzi in this regard is Nick Broomfield, who Bruzzi considers to be inherently performative (Bruzzi, 2000, p.198) in the manner in which he verbally interacts with his filmed participants and makes the film crew and apparatus visible in-shot. Nichols discusses Broomfield's processes as embodying the *participatory* mode, citing *Kurt and Courtney* (Broomfield, 1998) as a stark example of performance involving the filmmaker's intervention in the unfolding events he is documenting. In the film, Broomfield interrupts a ceremonial media gala to highlight Courtney Love's history of media intimidation (Love subsequently attempted to block the release of the film (Broomfield in Caldwell, 2015)). Such an act is highly reflexive in its negation of any disembodied objectivity and performative as Broomfield crosses the line between filmmaker and filmed subject.

There is a sociological component to the performances of Broomfield and other documentarians that have followed in his wake including Louis Theroux (E*xtreme Love - Dementia*, 2012), Morgan Spurlock (*SuperSize Me*, 2004), and Michael Moore (*Fahrenheit* 9/11, 2004). In this mode, audiences accept the documentarian as a component of the social world in which they interact, and consequently that the realities seen are necessarily *profilmic,* as the reality - of documenting reality. Sociological frames, such as Goffman's *dramaturgy* (Goffman, 1959) offer a suitable lens in this regard to consider nonfiction as one performance among many. His restatement of Park's "everyone is always and everywhere, more or less consciously playing a role…" (in Goffman, 1959, p.19) reflects the premise that social existence is a form of situationally determined self-representation. All social interactions are necessarily performed, both consciously and unconsciously. Added to this, all interactions are reflexive negotiations of our context, and the performances we adopt are always situationally motivated. It follows therefore that the production of nonfiction is itself a social situation that is negotiated and performed by all involved. The notion of documentary as a socially performed activity has gained currency in recent theory. Rogers (2015) expands on Goffman's dramaturgy in his critique of *Grey Gardens* (Grey Gardens, 1975). The main subjects, Edith, and Edie Bouvier Beale, once part of the New York elite, are filmed in a state of social decline, threatened with eviction from their once opulent residence now in considerable disrepair. In their circumstances of crisis, they are swapping between various performative 'masks' as they enact presentations of social class and femininity. Performance in *Grey Gardens*, for Rogers entails the films makers, who by their own covert

interventions instantiate these identities in the subjects (Rogers, 2015). Although they are rarely seen or heard, the filmmakers and their camera are clearly the recipients of Edie's performances.

Figure 1: Still from Grey Gardens, featuring 'little' Edie Beale (Grey Gardens, 1975)

Goffman's dramaturgy also makes a distinction between performative 'regions' (Goffman, 1959, p.106) where some performances are more public and 'on stage', and other more private, moments are 'back stage'. Menand articulates documentary's striving to make the off-stage – on-stage: *"*the documentary impulse is to catch the off-camera on-camera. The documentarian "sees" what is not supposed to be seen*,* the performance that Goffman called the backstage (Menand, 2009, p.299). Waugh frames this slightly differently, where subjects are either *representational* and performing-*for*, or *presentational* and performing-*to* the camera (Waugh, 2011). Representational acts appear to happen in and of themselves, without the performer's explicit acknowledgement of the camera (and by extension, the filmmaker, and audience). Conversely, presentational acts involve a conscious engagement with the filmmaker. In Goffman's terms, this could be described as a *front-stage* performance, where the performer is actively engaged in self-presentation and endeavouring to control the impression that is formed of them by the audience. The documentary instinct, as Menand states will be to blend the off and on-stage. To follow the dramaturgical metaphor, Edie Bouvier Beale can be said to enact a 'front-stage' presentational performance, that is poignant in her frequent 'over-acting'; moments when her performative mask slips.

Elizabeth Marquis (2014) conceptualises and systematises the dramaturgical aspect of documentary production, positing a framework of three performativity 'tiers' (p. 45):

1. *Tier one* comprises the general "everyday performance activity", as laid out in Goffman's theory.

2. *Tier two* contextualises tier one as a facet of documentary production; another social situation with its own situational demands ("the impact of the camera" (ibid.)) that must be negotiated by filmed subjects.

3. *Tier three*, "the influence of specific documentary film frameworks", concerns the ways that a specific mode of representation, or 'style' of documentary will inculcate a particular performance mode in the subject.

For Marquis, the second two *mediated* tiers provide the unique profilmic, situational context for the ever-present first tier. A feature of the framework is the progressive reduction of the subject's 'semiotic independence' when considered across the three tiers (Marquis, 2014, p.50). Their performances are constrained and expanded depending on the conditions set by the medium, the style of documentary and the creative decisions made by the filmmaker. The filmmaker for example may select the location within which the performance occurs, thus immediately re-contextualizing and impacting the subject and their performative 'work'. Lens angles are another example of this reduction in semiotic independence. For example, their performance may be represented, or presented (as per Menand) in interviews by either addressing the camera or addressing the filmmaker off-camera, as well as through enacted scenes, often used to provide 'colour' for the interviews through cutaway shots. When all tiers are considered, the subject may speak ostensibly for themselves, but their voice becomes modified in real time as a negotiation of the profilmic situation, and in the audio-visual context that the filmmaker presents it.

2.3.5 Camera Technology and Discourse

At the time of CVR's inception, comparisons were made with earlier observational movements regarding transparent 'fly on the wall' representations (Uricchio et al. 2016, 15). Comparisons such as this can be considered within the underlying remediative logic of CVR, as a means of generating transparent perspectives on reality, where the mechanics of production (and thus the author) are erased. The observational subgenre *direct cinema* endeavoured to erase the filmmaker to present a seemingly unmediated experience of reality; the CVR camera 'disappears' perceptually in 360° viewing, and the filmmaker also commonly erases their own presence by vacating the scene. Comparisons such as this are problematic however, as the similarities belie the very different ways that the viewer becomes embodied. Observational and CVR films are two instances of texts that use innovative camera technologies as the central feature of their discourse production. Films such as *Sex: My British Job* (Sex: My British Job, 2013a) by Nick Broomfield, and *Leviathan* (Castaing-Taylor and Paravel, 2012) are other useful comparisons regarding filmic perspective and ideologies of transparency. In Broomfield's case a hidden

camera is worn by the subject of the film, and in Castaing-Taylor and Paravel's case, multiple action cameras ('GoPros') are attached to the hull of a ship. I will begin with an account of the observational movements and follow that with discussions of *Sex: My British Job* and *Leviathan*, making comparisons with CVR throughout.

The observational forms of *cinema verité* and direct-cinema[5] signaled a move towards ethnographic, naturalistic modes of representation (Henley, 2010). The film theorist William Rothman describes a technologically enabled mobility affording new kinds of access and intimacy between filmmaker and filmed subject.

> "…a method of making documentary films, in which a small crew (often a camera person and sound recordist, sometimes only a solitary filmmaker) goes out in the "real world" with portable sync sound equipment and films people going about their lives, not acting"

> (Rothman, 2004, p.281)

The people 'going about their lives' were encountered in their private domain, such as John F. Kennedy in *Primary* (Primary, 1960) and the Bible salesmen in *Salesman* (Salesman, 1969). Observational films presented their subjects in 'behind the scenes' arrangements, which were combined with more outward presentational performances. While the observational forms were touted by proponents as being more *real*, the irony, as noted by Nichols (1983) and Rothman (2004) is that the forms inherit the sense of immediacy and intimacy from the aesthetics of fictional film: "classical-cinema's great stake in the realm of privacy and … of the everyday, the ordinary" (Rothman, 2004, p.286). For Nichols, it is the capacity in both to claim 'transparency' and to "capture people in action [...] letting the viewer come to the conclusions about them unaided by implicit or explicit commentary" (Nichols 1983 p. 17).

Verité had a clear visual style (Hall, 1991) but in its claims to realism and its fusion of fiction's aesthetics, it ran into ideological accusations of conceit. The American form, *direct-cinema* was derided in particular as naive from the outset for its claim to a transparent mediation of reality (Waugh, 2011). The strive for transparency bore a more sinister dimension where the viewer, absorbed by the directness of the form, would be more open to suggestion. The form was thus "denounced as a transparent purveyor of ideologies" (Hall, 1991, p.26). In this light, Noel Carroll quipped of Verité that "it opened a can of worms and then got eaten by them" (in Carroll Kahana

---

[5] Also referred to as 'American verité' with both movements collectively referred to at times as 'verité'.

and Musser, 2016, p.506).

European verité took a more nuanced approach, that acknowledged reflexivity and performativity, thus circumventing some of the criticisms levelled at its American counterpart. Geiger points out that the reality claimed by Drew et.al. in direct cinema was simply not available to be filmed. The (filmed) 'other' is "always technically, temporally and spatially removed from the filmed event and its moment of occurrence" (Geiger, 2008, p.4). Any 'actuality' is therefore always pro-filmic and thus altered through the practice of representation. Regarding immediacy and the invisibility of the filmmaker, Geiger suggests that "where directorial control is most invisible, absence merely masks a ghost" (ibid, p.6). Filmmakers such as Jean Rouch worked under this assumption and took their cues from the reflexive nature of ethnography, forming a more nuanced view of authorship. In *Moi Un Noir* (Rouch, 1958), Rouch's own voice is interspersed between the scripted performances of the film's subjects. While Rouch's voice is heard, any notion of his didactically 'transmitting' anything to the audience is negated by the nature of the performances. Compared to direct cinema, Rouch negotiates the connection between viewer and subject, with very different results. Geiger (2008) describes a scene from *Moi, Un Noir*, questionable by today's standards, involving the not-visible character Ganda's amorous voice-over, transposed onto the silently undressing character Dorothy Lamour. A composite view forms; of Ganda's, the filmmaker, and viewer, all in a 'private' shared gaze of Lamour. In the construction, Rouch is attempting to create a shared cross-cultural experience in which he and his camera are directly implicated, highlighting his own, his subjects, and the viewer's shared, sexual humanity*.*

Figure 2: Dorothy Lamour (Moi, Un Noir, 57:23)

Similarities between verité's erasure and CVR's break down when one considers the nature of embodiment afforded by the different cameras. Unlike CVR, the author's position in conventional filming is signified both literally and figuratively through the camera's positioning and

movement, omnipresent in negotiating the psychological space between subject and spectator. Albright (2011) termed this notion 'cinematographic embodiment', the notion that there is always a physical instatement of the author, inferred through the captured movements of the mobilized camera. It is the surreptitious filmmaker, silent yet omnipresent that Geiger infers a *ghost*-like quality (2008), referred elsewhere as a 'camouflaged' entity (Waugh, 2011, p.84). This highlights an important point: in conventional filming, there is almost always a *person* behind the camera that is engaged in the social activity of filming and who is jointly responsible for the resulting performances. The CVR camera's lack of a 'behind' presents a sharp divergence in documentary practices, which is perhaps missing in theoretical discourse in favour of the spectator's immersed social experience of viewing.

Cinematic embodiment is reformulated in Broomfield's film *Sex: My British Job* (Sex: My British Job, 2013a) where embodied performance is a technical feature of the film. The film documents the female journalist Hsiao-Hung Pai's experience as she goes undercover, posing as a cleaner in an illegal brothel. Pai documents the experience via a hidden camera embedded in her spectacles and the film is viewed by us as her first-person, embodied experience and gaze.[6] 'Cinematographic embodiment' is reformulated here, as the author, subject and viewer become fused in Pai's acts. Transparency, immediacy, reflexivity, and performance are reconfigured in this curious mix. *Sex: My British Job* continues the legacy of observational cinema in its use of the mobile camera in private spaces. The technology here however allows for a very different kind of private space as the brothel workers are unaware of the camera. Pai's undercover intervention is *constative* as her presence generates a series of performances in those around her, principally Mary, the brothel owner whose own performance hides a particularly sinister agenda of inculcating Pai into prostitution. From a dramaturgical perspective, the film presents a paradoxical overlapping yet mutually exclusive set of situations. Where Pai and Mary are in a filmed exchange, Pai is negotiating a situation that is in large part a filmic situation; Mary's performance on the other hand, and the mask she wears in her act are the service of a different situation entirely.

---

[6] Her filming is also interspersed with the brief secret exchanges with Broomfield which are filmed by a third-party camera operator in 'normal' filming mode. In this regard, Broomfield maintains an equal or higher-level of authorship (and perspective) to Pai depending on how he is considered vis a vis their collaboration.
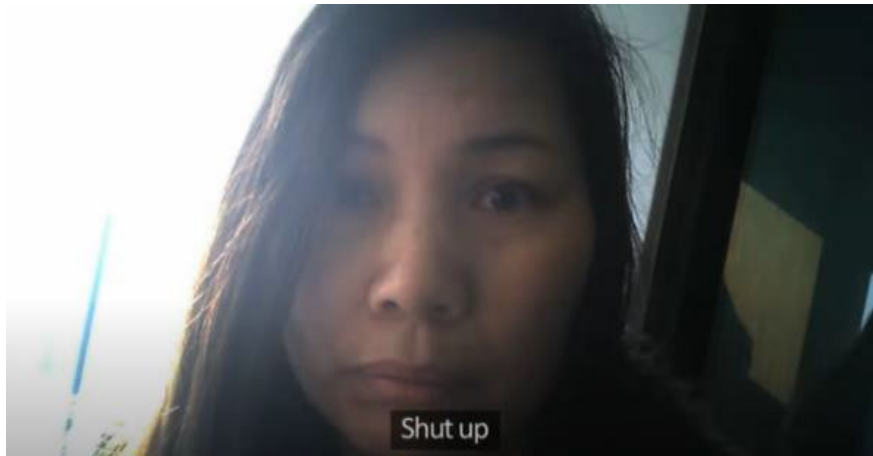
Figure 3: Screen capture from Sex: My British Job. (Sex: My British Job, 2013b).

Leviathan (Castaing-Taylor and Paravel, 2012) operates on a different premise to *Sex: My British Job.* Rather than centring its representations from a singular point of view, the filmmakers decentre and fragment the filming processes by placing multiple action cameras, in, under, and around a shipping trawler at sea, often in parts inaccessible to conventional cameras and their operators. In doing so they set up a novel engagement with the trawler which, like Pai, becomes both the film's subject and point of view.

Figure 4: Screen capture from Leviathan. (Castaing-Taylor and Paravel, 2012)

Leviathan's directors Castaing-Taylor and Paravel consider the decentring of image capture as fundamental to the film's mode of address:

"Looking at our footage, we were struck by a paradox, which was that we felt we were seeing moving images and sounds that were simultaneously divorced from shoulder-mounted, optical POV that you associate with documentaries and in particular with

non-fiction cinema; it seemed to be completely disembodied, and separated from directorial intentionality in that way. And yet it seemed to be much more yoked to a subjective, embodied experience of the world that you would have when you're actually in the world, not when you're just making a film. So it was both disembodied and embodied, and sort of an objective manifestation of a subjective experience."

(quoted in Jaremko-Greenwold, 2013)

For the filmmakers, an unexpected immediacy resulted from the removal of the physical human author from behind the camera, as a cinematographic 'dis-embodiment' but also a re-embodiment of sorts as a direct contact with the kinetic physicality of the ship and sea. As a *sensory ethnography film* (SEF) developed within the Sensory Ethnographic Lab in Harvard, *Leviathan*'s producers are operating from a defined ideological standpoint that favours the reinstatement of sensory experience as a site of understanding. Pavsek (2015, p.4) describes the film as typical of SEF in rejecting the 'logocentrism' of nonfiction film that privileges "propositional knowledge, narrative voice over, the pre-researched and the pre-textualized […] false clarity of explanation, and didacticism more generally" (ibid page 5), with particular ire for subjective narrative-driven tropes that have become the mainstay of documentary film (ibid page 4). In this regard SEF can be described as going against the grain of much of nonfiction's evolved modes discussed earlier that favour subjective and reflexive modes of representation.

Regarding the use of the cameras specifically, Castaing-Taylor and Paravel assert that the autonomous visuality that *Leviathan* affords, constitutes discourse in and of itself, where the image is otherwise generally relegated to that of *illustration* (ibid.) and where the predominantly language-as-discourse paradigm subordinates the image to language (in Steinberg 2015, 81). The mode of address is, for Castaing-Taylor and Paravel, of raw 'sensory input' intended to bring about the immediate "'affective and embodied' aspects of 'social existence and subjectivity'" (Pavsek, 2015, p.5). Waugh's ideological critiques of transparent media echo here, as *Leviathan's* makers exclude the extrinsic factors involved in the production, from the choices of the technological apparatus of filming, to the placement of the camera. To this end, Pavsek (2015) comments that in SEL, the embodied viewer is no less subject to "embodied oppression […], one perhaps as ominous as that of the sorts of conceptual domination that an authoritative voice-over might impose" (Pavsek 2015 page 7).

Like the hidden camera in *Sex: My British Job*, and the fragmented camera array in *Leviathan*, the CVR camera alters the dynamic between the embodied camera, the filmmaker, and the spectator.

What is often missing in discourse surrounding CVR is its 'hands-off' production mode. For example, a moving CVR camera tends to be moved with flying drones, as they provide minimal intrusion in the shot and can provide smooth arcs of movement. Thus, the CVR apparatus resists the kind of embodied mobility described in observationalism, as well as *Sex: My British Job*. There are exceptions in CVR films such as *The Protectors* (Kathryn Bigelow 2017) which uses participants visible in the film as de facto camera operators who must surreptitiously carry the CVR camera (and the spectator) with them as they move through the scene. *The Protectors* is an obvious example of the clunky contrivances required to manually move the CVR camera (as shown in Figure 5, below). In any event the omnidirectional nature of CVR camera lenses will betray the artifices used to add camera movement. On a deeper level, the embodiment enabled by the mobile (conventional) camera in classical documentary, as well as the novel embodiment afforded in *Sex: My British Job*, is fundamentally at odds with CVR's perspective which is always at least partly the spectator's own reflexive embodiment.

Figure 5: The CVR camera mobilized by a visible participant in The Protectors (3:36)

Leviathan points to other limitations of CVR regarding non-human perspectives. The sensory ethnographic impetus lies in its decoupling of the viewers perspective from the embodied authorial position, presenting instead a fragmented series of bespoke views which in their totality constitute *the ship.* This is impossible in CVR, as the camera must inhabit a range of natural human heights, and to deviate from these heights will present a distorted sense of space and disorient the spectator (Keskinen et al., 2019).

2.3.6   Nonfiction Frames For CVR

Nonfiction CVR is critiqued by Kool (2016b) from an ethical perspective, who reflects on CVR's modes of production and claims to realism in social contexts. The key component for Kool is in

the invisibility of the filmmaker (2016, p.6) which engenders a sense of unmediated realism that is open to propagandistic uses (ibid.). Echoes of the observational movement are found in Kool's critique of Geiger's *ghost* (2008), and Waugh's *camouflaged* author (2011). The pattern throughout such critiques involves the surreptitious author, maneuvering themselves 'out of sight', figuratively in the case of the observational movement and literally in CVR's case. Kool quotes Arora, producer of *Clouds Over Sidra* (2015a) for his account of the production process where he must necessarily vacate the scene (O'Neill, 2015; Kool, 2016b) affording the process a 'purity' of sorts, a notion Kool dismantles. Essentially, the lack of framing in CVR creates the impression that 'all is there to see' but this belies the wider editorial decisions, including the choice of location and positioning of the camera (Kool, 2016). The CVR camera by its nature ostensibly loses its ability to exclude, but can instead editorialise by omission. As such 'framing' exists but in the broader figurative sense, of *framing the situation*. Finally, Kool points out that the 360° image is just one part of the overall nonfiction text, and its realism can allow other augmenting factors to influence the suggestive CVR spectator, essentially 'piggy-backing' on the image's realism to further achieve the aims of the film. He uses music as an example but stops short of discussing the tailored, strategic use of language in CVR, an aspect I address in this study (e.g., section 8.1.2 (Medium)).

As part of her wider analysis, Nash (2017) uses the United Nations VR output (United Nations, 2018) to describe the potentially problematic ethics regarding the immersed CVR spectator. Using *media witnessing* as a broad frame, Nash probes nonfiction CVR's capacity to engender empathic responses to the suffering of represented subjects. Taking Slater's *place illusion* (Slater, 2009) as the basis of the CVR's embodied realism, and the sense of presence that witnessing requires, Nash presents an issue of 'improper distance'. In mediations of disadvantaged subjects, *proper* distance ensures that we are brought close enough to experience and empathize with the other, while retaining sufficient critical distance to reflect constructively on their situation (as an 'imaginative engagement in the third person' (Nash, 2017, p.6)). The issue for Nash, is that by collapsing the distance between viewer and other, a spectacle emerges, and the 'directionality' of the exchange becomes less distinct. In 'being' the other for example (as first-person perspective), we may not necessarily take on the views and experiences of the other but instead imbue the situation more overtly with our own preconceptions. This idea of distance articulated by Nash has a semantic correlate of *semiotic distance*, which I will explore I detail in chapter 8, section 8.1.3. (Experiential Distance).

As part of her analysis of *Waves Of Grace* (Milk and Arora, 2015b), Nash points to CVR's

asynchronicity as an aspect of its (in)capacity to provide fulsome witnessing experience. She describes a scene involving the main character Decontee Davis in a moment of distress, which is captured over an extremely long shot duration (37 seconds). As the spectator has no technical agency in the experience (except by rotating their view), a paradoxical sense of closeness yet powerlessness results, which Nash compares to the 'uncanny valley' (Nash, 2017, p.10). Essentially, the scene is sufficiently real to feel contemporaneous, yet fundamentally unavailable, much like Barthe's 'illogical conjunction of here and then' (quoted in Metz, 1974, p.10).

2.3.7 Nonfiction: Concluding Remarks

Nonfiction's complex theoretical space contains a common technological and ideological thread that takes transparent immediacy as its dominant logic. CVR appears to recycle claims made in previous transparent forms, where the author is seen as a kind of interfering entity, and where a more direct knowledge of the world can be made available in their absence. I have described the observational movements, *Sex: My British Job*, and *Leviathan* as pointing to the desire to affect an immediacy, predicated on reducing the author's visible involvement in the film. Kool points to the problematic aspect of CVR being considered this way. Where broader editorialising decisions impact what is shown in the CVR documentary, the transparency it claims is reduced to an aesthetic, as was the case with the verité movements.

There is a schism and a tension between conceptions of 'reality' that *should* be able to be shown by technical means, and the many kinds of performances that *must* take place to show it in a meaningful fashion. Dramaturgy proves to be a suitable frame for describing nonfiction production in CVR, as it considers the process and the output of nonfiction texts in a holistic way, where author, subject and the means of production are unified in a singular social situation of discourse production.

Taken as a whole, this section outlines the many pertinent dimensions of this study and adds context to the social semiotic nature of the methodologies used. Nonfiction genres, as described here reflect the technological and social-performative aspects of discourse production. When considered from a linguistic perspective, this will inform the study of language-use, as the product of profilmic situations, that are informed by and constrained by genre.

2.4 Concluding Remarks

This chapter explored CVR from a technological, formal, and nonfiction perspectives. Genealogical

theories were presented as a suitable means of characterising CVR at a high level, in its current state of evolution. Remediation was introduced as a conceptual frame for considering CVR's hybrid status, allowing for a structured exploration of CVR as the confluence of other media forms, namely photography, film, VR, and immersive theatre. Remediation also allowed for CVR to be described for its ontological makeup, and its *transparent-immediacy* logic.

As an exploration of CVR's formal qualities, the chapter provides context for specific aspects of the study's methods. Accounting for the similarities and differences between CVR's constituent parts aids in the differentiating of CVR as a semiotic technology. For example, Barthes observations of the photograph, as the 'illogical conjunction of here and then' resonate with CVR's MODE, and the sensory and spatiotemporal contact afforded to the CVR spectator, who are themselves reflexively immersed in CVR's 'here and then'.

Where nonfiction is concerned, the chapter provided a grounding in genre and the social goals that nonfiction texts aim to address. Nonfiction is inherently complex and at a times paradoxical, and I have shown the various ways that these inherent complexities are framed ideologically and how they play out in nonfiction praxis. By foregrounding the camera as a technological instrument and proxy for the viewer's experience of the social world, I have contextualised the perspectives CVR *purports* to give and the kinds that it is *capable of providing*. This is invaluable when considering the practical implications for CVR producers when the technological act of filming constitutes the social context of meaning production.

3. The Systemic Functional Frame

This chapter provides the theoretical frame for the study and grounds the methods described in the next chapter. The multimodal analytical methods used in this study are an offshoot of the *Systemic Functional Linguistic* framework (SFL) developed by Michael Halliday in his seminal work *Language As Social Semiotic* (1978). SFL is a metatheory of language, with an emphasis on the social production of meaning. SFMDA[7] theories adopt the SFL conceptual framework and make it practicable in non-linguistic and multimodal texts. In this chapter, I aim to unpack SFMDA in a 'root and branch' sense: its roots being SFL; its branches the diverse SFMDA frameworks that have adapted SFL for use in multimodal texts. I will outline the SFMDA theories, systems, and semiotic resources currently available in the analysis of image, sound, text, and other modalities, and will present some additions from related fields.

The chapter begins with a description of SFL's core tenets including the stratified discourse model on which this study is based. This is followed by an account of *functional variation*: discourse as the realisation of universal social functions. I will enumerate the visual semiotic resources associated with *ideational*, *interpersonal,* and *textual* meanings and suggest how some of these resources can be augmented in the analysis of immersive texts. Following this I will describe discourse at the higher levels of context, where discourse realises *situational* variation. This will include a breakdown of *register,* the system used in analysing the contextual aspects of language- use, and a brief discussion of *genre*, the highest level of discourse in the systemic functional model.

The latter part of the chapter addresses the concept of *modes* as units of multimodal meaning, and as the units of analysis in this study. Modes are considered from several theoretical perspectives, and I present a working definition for this study. Following this I will address intersemiotic meaning and the nature of modes in context. I will discuss several ways that meaning is created multimodally, including functional organisation, cohesion, and conjunction.

---

[7] The term 'multimodal' can refer to any study that uses image, text, action and other modes as its raw data. This study adopts the framework of SFL and as such I will use the term Systemic Functional Multimodal Discourse Analysis (SFMDA) popularized by O'Halloran (O'Halloran, 2008a).

3.1 SFL and SFMDA as Social Semiotics

The principle of social semiotics is summed up by Halliday in his seminal work *Language As A Social Semiotic* (1978) where "…there can be no social man without language and no language without social man" (p.12)*.* Following the works of his mentor Firth (cf. Love, 2012)*,* Halliday described language and society as 'a unified conception' where one cannot be extricated from and studied without the other. Where language was previously considered for its psychological effects on the individual, the individual was now considered to be *using* language as "the means whereby the various social relationships into which he enters are established, developed and maintained" (p.16). Semiotics are 'social' because meaning-making is reflexive in its social environment of use. The 'means' are the semiotic resources available to those involved in language acts, a concept extended in SFMDA to the wider sense of multimedia expression. Semiotically, the foregrounding of 'semiotic resources' in SFMDA is significant in that it signals a move from the more immutable *sign*, in the Saussurian, structuralist sense (van Leeuwen, 2005).

The *functional* aspect of SFL and SFMDA relates to the underlying functions that language serves in any moment of successful discourse. Referred to as the *metafunctions*, they are the *ideational*, *interpersonal*, and *textual* functions which relate to the construal of experience and logic, the social relationships between interlocutors, and the organisation of information in language exchanges, respectively. The *systems* are the systems of functionally specific choices available to producers and receivers of discourse. SFMDA takes as its main premise that these metafunctions are sufficiently universal to be applied to any kind of media text. As a functional analysis of *CVR* texts, the study addresses the specific choices afforded in CVR discourse by the 360° camera, and where other choices typically available in other mediated discourse become constrained.

The emergent nature of CVR is also compatible with the SFL/SFMDA framework which considers language as constantly evolving, from the moment-by-moment evolution of discourse between interlocutors, to the longer-term cultural adaptations to language-use. In short, SFMDA provides CVR analyses with solid theoretical grounding and implementable systems of analysis, systems that are open and malleable enough for CVR analysis to contribute back into the SFMDA field.

3.2 The stratified modelling of language

SFL adopts the principal that context always manifests in language. This stems from the earlier

work of Firth, who emphasised the situationally determined aspect of speech.

> "What is said by one man in a conversation prehends what the other man has said before and will say afterwards. It even prehends negatively everything that was not said but might have been said"

(Firth, in Love, 2012, p.36).

Language is both determined by the situation at hand, while also reflexively producing the situation, as discourse. SFL models this contextual nature of language as a stratal system, from the utterance to its situation of use to the abstract cultural and ideological context that shape the situation. Models vary but essentially these strata are *language*, *context*, and *genre* (Figure 6, below shows these strata but includes genre in context).
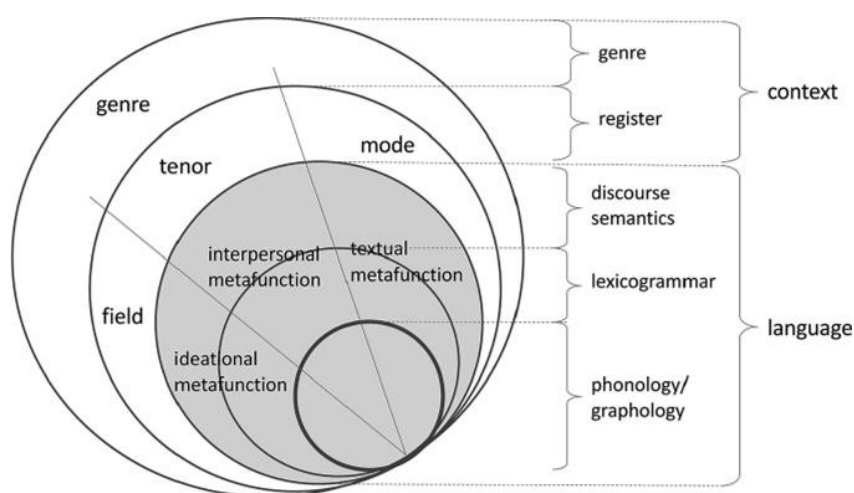
Figure 6: SFL's stratification model, Hood and Hao (2021)

The language stratum is systematised through the *metafunctions*, outlined above. Situation is systemised as *register*, where instances of language are characterised for the way they reflect their context of use (e.g., playing a card game or telling a story). Each situation in turn realises deeper socio-cultural norms. This is codified in SFL's model as *genre*, where discourse occurs as *social processes* (Martin and Rose, 2008, p.6). The underlying generic processes (e.g., sharing one's personal experience in an *anecdote*, or arguing a logical point in a *debate*) are considered to manifest implicit ideology (Martin, 1992, p.121).

The stratified system can be perceived as operating in two directions. From *above*, genre determines the 'rules of engagement' for a given discourse situation, influencing the lower-level choices made in language. From *below*, linguistic choices reveal the kind of situation at hand, and

how it interfaces with genre. This bidirectional realisation is termed *metaredundancy* (Lemke 1993) as each level redounds with the one above and below. Genre, register and language, while necessarily separated in the model are always co-occurring dimensions of a single instance of discourse (hence the enfolding circles seen in the illustration of the model).

Martin (1992; 2002) further adapted Halliday's original modelling of the language stratum with implications for multimodality. The language stratum, in its original Hallidayan conception follows Hjelmslev's concept of the sign comprising *expression* and *content* (Hjelmslev, 1961). In SFL, the expression plane equates to the printed word or the phonetics of speech, whereas the content plane realises the semantic aspect of language. Martin (1992; 2002) further subdivided the content plane to include an additional *discourse semantics* plane, where language could be realised on a more abstract semantic level. The impetus in doing this lay partly in the fact that the lower-level lexicogrammar can say different things that essentially mean the same thing (Martin, 1992). His examples from the *Hitchhiker's Guide* demonstrate the point where the clauses 1) *Ford is smiling because Trillian arrived,* and 2) *It pleases Ford that Trillian has arrived,* present different lexicogrammatic meanings, while essentially meaning the same thing. Example one denotes a *behaviour* ('smiling') and example two, a thought ('pleased') yet they both encode *Ford's pleasure at Trillian's arrival.* The discourse semantics plane is useful in analysing larger tracts of text, but its relevance for multimodality lies in the abstracting of the language plane into *discourse units* (Martin, 1992) that are realised by, but not 'tightly bound' to the underlying lexicogrammar. In this regard, Martin's discourse stratum is a conceptual bridge between language and other modes of expression. Without it, images would otherwise require 'lexicalizing' for their meanings to be derived systemically. The other significant aspect of Martin's discourse semantic stratum is how it caters for flexible analysis of large tracts of text. For this reason, film semioticians have used his discourse stratum as the framework for analysing shot sequences in films. Tseng (2013), Bateman and Schmidt (2013) and van Leeuwen (1991) use the concept of the semantic discourse stratum and its units as the basis of their film analyses, as well as comic strips (Lim Fei, 2013). It is in this light that Martin's framework is useful in this study, as it provides generalised semantic system of discourse units that can be employed across multimodal phenomena, as they arise in the CVR texts.

3.3 Functional Variation in Language

Regardless of how or where language is used (speaking to a friend, writing a book, taking a photograph) language-users in the production of discourse are fulfilling the three fundamental functions of language (adapted from Andersen (2017)):

- The *ideational metafunction*: constru[ing] our experience of both the inner (mental) and the external (social and physical) world
  - o *The experiential function*: construing experience, as entities involved in processes
  - o *The logical function*: the connection between events [construing] meaning in a more abstract way
- The *interpersonal* function: the interaction between the producer and the perceiver (of a text)
- The *textual function*: the organization of the resources used to create cohesive and context sensitive texts

Functional meaning is found at the language level by analysing the ways that the metafunctions are realised in the production of text. Each function comprises a system of options available to producers, as the semiotic resources available in discourse. For the most-part, these resources were developed in SFL and have been adapted for non-linguistic and multimodal contexts. As such, further discussion of the metafunctions will include multimodal resources that I consider apposite to studying CVR's modality, as the immersive combination of spatial, kinetic, gaze and diegetic sound, as well as voice and music. Taking each metafunction separately, I will begin by describing how SFL systematises each metafunction, followed by the SFMDA equivalents.

3.4 The Ideational Metafunction

The *experiential* metafunction comprises the system of *transitivity*, the relations of people, enacting processes, in circumstances (e.g., 'the dog was barking in the garden'). The *logical* metafunction, through the logico-semantic system, relates experiential meanings to each other to make logical meanings (e.g., 'the dog was barking in the garden, *so we left'*). This section looks primarily at the former, experiential dimension, and the logical function is discussed later as a property of intersemiosis and conjunction.

Transitivity in SFL comprises different process types such as *material processes* that involve multiple participants as the 'doers' and the 'done-to' (actor and goal respectively). Other processes involve *behaviour*, *thought*, *speech*, *being*, and also *relational* processes that relate qualities and identities to participants. Kress and van Leeuwen adopt the term 'representational' for experiential visual meaning where "Any semiotic mode has to be able to represent aspects of the world as it is experienced by humans" (Kress and van Leeuwen, 2006, p.41). In their seminal work *Reading Images: The Grammar of Visual Design* (Kress and van Leeuwen, 1996)

they systematise visual transitivity (as well as other functional meanings), the results of which have become canon for visual SFMDA analyses. Language and images encode processes in different ways, for example a person cannot be explicitly visualised as *thinking* or *feeling*. As such, unlike SFL, visual grammar operates on a limited distinction of *narrative* and *conceptual* processes. Narrative processes reflect visible activity with vectors emanating from and/or connecting figures. These include a variety of types depending on whether the action is proactive or reactive and how many participants are involved. Where vectors are entirely absent, processes are *conceptual* (p.50), and objects and people are instead construed taxonomically for what they represent. These kinds of processes require reimagining in CVR where conceptual imagery, as typically described in visual grammar, involve the extracting of entities from their natural surrounds to display their essences. CVR's 360° images, being omnidirectional, make this difficult if not impossible where visual settings are always available.

3.5   The Interpersonal Metafunction

Where the ideational function concerns 'language as reflection', the interpersonal function is concerned with 'language of action' (Halliday and Matthiessen, 2013, p.30). In this regard, the interpersonal function deals with how interlocutors enact their relationships through speech, thus managing the social aspect of the interaction. The question 'would you like an apple?', and the order to 'take this apple!' each reflect the same meaning yet each realise a very different set of roles in the interaction (asking and insisting). In this example, the text 'project[s] the relation' (Kress and van Leeuwen, 2006, p.41).

The primary interpersonal systems of language are *MOOD* and *modality*. MOOD allows interactants to choose whether to *ask, request, state, offer*, or *make commands and demands,* and in doing so, to establish a social role in the interaction. MOOD has been extended into multimodal theory in different ways. Martinec (2001) established a non-linguistic MOOD system realised in embodied action where for example an *offer* (of goods and services) can be coded indexically, where goods can be literally proffered to the receiver (a 'take this from me' hands gesture). Culturally specific 'statements' can be made as physical gesture, such as the two-fingered (index and little finger) gesture 'you are a cuckhold' (2001, p.132). Where Martinec's action schema makes a valuable step from language to embodied meanings, they are largely paralinguistic and lexicalizing actions. Kress and van Leeuwen's system of *contact* provides a more fundamental reformulation of MOOD. As with their transactional systems, the interpersonal system of *contact* contains fewer variables than are available in language, namely *offer* and *demand*. A demand image, would involve a represented participant directing their gaze to the camera-viewer,

thus constituting an 'image act' approximating to Halliday's asking/requesting/demanding, essentially enjoining the viewer to act (Kress and van Leeuwen, 2006, p.118). Conversely, an image that provides only a bystander perspective on events portrayed, reflects a contact of *offer,* where the viewer remains a more neutral observer (p.120).

Modality in the linguistic sense is concerned with a speaker's stance. A statement is *modalized* when it conveys usuality and probability (Eggins, 2004, p.172). To say that I will 'probably', 'possibly' or 'definitely' do something reflects my confidence in the statement and what it represents. Modality in visual grammar differs significantly as more of a global property of the image, rather than any singular part therein. Kress and van Leeuwen measure modality as an image's *naturalistic coding orientation* where for example colour and style denote the fidelity available to the viewer, in the image (Kress and van Leeuwen, 2006, p.166). The photographic image is the prototypically 'faithful' modality, albeit only as we are culturally conditioned to think so (Kress and van Leeuwen, 2006, p.156). Visual modality in itself is not considered in detail in this study as the photographic modality was itself a sampling criterion, setting the boundaries of the phenomenon being studied (highly realistic, photographic immersive texts). Modality does have resonances intersemiotically however, where music has the capacity to augment the naturalistic modality of a text (cf. Wingstedt, Brändström and Berg, 2010).

The omnidirectional CVR camera has implications for the closely related visual systems of *social distance* and *involvement.* In SFL, social distance is primarily a property of register as 'distance' is characterised across the text as formality, familiarity and friendliness (Mechura, 2005). Unlike spoken language however, distance can be explicitly realised visually, equating visual proximity as semiotic distance (cf. Lam, 2016 for typology of distance). The camera-as-proxy relationship discussed in the previous chapter (section 2.2.2) enables interpersonal grammar where perceived distance between the subject and the viewer suggest 'closeness', based on Hall's *proxemics* (1966) (Martinec, 2001; Baldry and Thibault, 2006; Kress and van Leeuwen, 2006). Hall established a set of spatial distances, each affording the perceiver a degree of visual, aural, tactile, and other informational options that are specific to kinds of relationships from *intimate*, through *personal*, *social*, to *public* (2006, p.124). Distance for Kress and van Leeuwen equates to shot sizes that synthesize 'biographies' between those represented and the spectator (Ibid.). The perception of distance is augmented by the angles that people take up relative to each other. The way one orients towards or away from another is systemized as *involvement.* While both are usually considered together, Martinec (2001) formally combined proxemics and involvement in a single taxonomy of *engagement* (p.120). For example, six inches is undeniably intimate when the interactants are in frontal involvement but at a side angle, the engagement

becomes more personal.

Involvement also overlaps with contact and perspective, where a frontal gaze 'confronts' the viewer through maximal involvement (as demand), and where an oblique horizontal angle places the observer at a degree removed (as offer). In this regard, involvement also realises bystander and first-person perspectives. Once the subject reaches an oblique angle, they become psychologically detached from the represented subject and as the angle continues to increase and begin to share the space and activities of those represented: a shared perspective from 'behind' the subject so to speak.

A conceptual leap is required when considering social distance and involvement in CVR. CVR images present distance differently to conventional imagery where it is the *size of the frame* that denotes closeness (Kress and van Leeuwen, 2006, p.124). How this framing is achieved is arbitrary; the photographer can be physically close to the subject, they might be far away yet 'zoomed in', or they may crop the image after the fact to achieve the effect of closeness. In CVR closeness is an *index of the actual distance* between the camera (and consequently the spectator) and the filmed subject. Also, the omnidirectional nature of the lens means that where the spectator must be positioned relative to more than one participant, standing *with* one, and *away* from another for example. CVR becomes a distance-matrix where the spectator negotiates their 'place' in all directions simultaneously. A curious aspect of CVR specifically regarding involvement is where the spectator may themselves be angled toward or away from a filmed subject, where otherwise a film viewer would always be assumed to be facing the image. Adam Kendon (1967; 1990; 2010) is apt in this regard in extending Hall's proxemics schema, specifically through his *use-space* and *formations*.

3.5.1   Kendon's Socio-Spatial Systems

Kendon developed a series of cross-cultural socio-spatial models that are perceived more readily in CVR than in framed images. They comprise the overlapping aspects of use-space and formations (Kendon, 2010). Use-space is concerned with how the environment is dynamically structured into *transactional-segments* by ongoing activity. The use-space is exclusive to the activity as it enfolds: where an actor needs to fulfil some transactional goal, their doing so will delineate their transactional segment from the surrounding space. Formations are more socially oriented as "spatial-orientational arrangements sustained over time [...] through the cooperation of the participants" (Kendon, 2010, p.5). Formations are the shapes that people make in establishing usable zones of engagement where conversations can be enacted with ease. Formations by their nature include ratified participants (the 'in group') delineated from

those not yet 'approved' for inclusion. A prototypical formation is an *F-formation* comprised of a series of concentric zones, as shown in Figure 7, below.
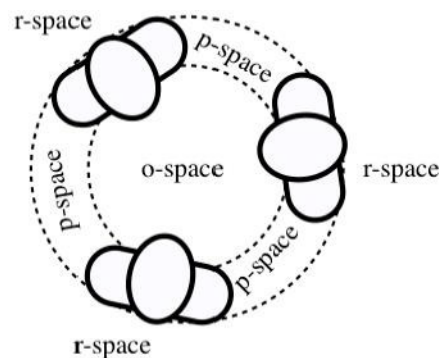
Figure 7: An F-formation (Marshall, Rogers and Pantidi, 2011)

The *O-space* is the core of the formation that exerts a kind of social gravity on the formation's participants who actively sustain and protect it (ibid. p.5). The *P-space* comprises the participants themselves as an inward-facing barrier of sorts demarcating the o-space from those who are not yet included. Beyond this is the *R-space*, a buffer zone where only those who wish to become included will enter. This zone is highly charged and typically avoided by those passing, who will physically divert their path while avoiding the gaze of those in the P-space (who will reciprocate by ignoring). If an 'outsider' is making a bid for entry into to the formation, they will enter the R- space and wait to be admitted, at which point those in P-space will re-orient themselves accordingly (or embarrass the bidder by remaining closed to their entry).

Use-space and formations are effective resources for describing the novelty of being immersed in social fields of activity. Use-space allows for transitivity to carve up the space meaningfully, and F-formations as highly charged cross-cultural social phenomena have ramifications for the immersed 'social spectator' in CVR. The placement of the camera will at times position the spectator in and around these resources, adding semantically to the resources of distance and involvement.

3.6   The Textual Metafunction

Most image producers, down to the casual photographer have a rudimentary, if unconscious awareness of information structuring through basic principles such as framing-out unwanted elements and foregrounding important things. The textual metafunction is concerned with this kind of semiotic choice which structures and organizes information in texts. As purely

structural, the metafunction is the means through which the ideational and interpersonal become 'enabled' in "the ongoing creation of a semiotic realm of reality" (Halliday and Matthiessen, 1999, p.8). As with the other functions, SFL provides the linguistic premises which have been extended in multimodality, most notably through the *compositional* systems developed in visual grammar (Kress and van Leeuwen, 2006). CVR, lacking the rectangular bounds of the conventional image must organise its visual information differently in the 360° space. The textual, as an organising function, also encompasses other functional concepts relating to cohesion and Intersemiosis, where information is structured between modes and across time. I will provide a grounding in the metafunction here and expand on the concepts in the later subsection dealing with intersemiosis (sections 3.9 and 3.10).

The linguistic structuring of language is organised in SFL largely through the systems of *theme, rheme, given* and *new.* The theme portion of a clause reflects the angle of the information to which the rest of the clause (as rheme) relates. By alternating what is placed as theme the angle of a clause changes and thus the information is framed differently. All that follows is rheme, where new elements are introduced. The following are examples of alternating theme (underlined), where the angle, and thus the meaning of the clause is changed.

1. <u>George</u> brought his *dog* to the vet
2. <u>The dog</u> was brought to the *vet*

Martin (1992), and Martin and Rose (2007; 2008) developed the system of *periodicity* to describe this patterned flow of information across written and spoken text. The structured flow of information across time has been described as a *rhythmic* feature of written language (Martin, 1992), in the phonology of speech (Martin and White, 2005) and also in multimodal texts (Martinec, 2000; van Leeuwen, 2005). Rhythmically speaking, where theme is used to provide an angle on the clause, it also functions to demarcate a wave of information that will 'crest' at moments of 'textual prominence' (Martin and Rose, 2008, p.34). Where the theme's prominence as angle was noted above, it also frequently constitutes what is already known and thus *given*. The other point of textual prominence will be the part of the clause that initiates *new* information (italicized in the examples above). This new information will resolve the theme and set up the next wave of information thereafter. In speech, the new element will be heard phonologically as the onset of the final tone group with prominence given to the information to be assigned 'newsworthy' status. In the first example given above, the *new* components could be the 'dog' or 'vet' depending on the phonology and thrust of the statement.

45

Visual grammar's systems of *composition* are described as integration principles whereby "Integration codes serve to produce text, to place the meaningful elements into the whole, and to provide coherence and ordering among them" (Kress and van Leeuwen, 2006, p.201). The compositional system is comprised of the subsystems: *salience, Information value,* and *framing* (ibid. 177). These systems reformulate the textual function for visual media while retaining the essence of Martin's linguistic 'moments of textual prominence', realised spatially as opposed to sequentially in language.

Visual salience approximates loosely to *theme* in SFL (Baldry and Thibault, 2006). As a figure-ground construct, salient elements are textually prominent by their being *variant* and "substantial and distinct with respect to their background" (Ibid. 2006, p.199). Salience can manifest in different ways: *large against small*; *red against green* or any other optical figure-ground relationship. *Movement against stillness* provides the most stark kind of salience, as borne out in studies of visual perception (Itti, 2005). Baldry and Thibault (2006), make a direct comparison between salience and theme in their analysis of shot sequences in television advertisements, where salience is contributing to "a higher order visual *thematic* system" (p. 188 my emphasis). Kress and van Leeuwen describe salient objects as having 'visual weight'. VR studies discuss salience indirectly as 'points of interest' (POI) (Brillhart, 2018a) frequently used to direct the spectator's engagement around the 360 space. In this light salience and POIs are discussed later as co-extensive in the context of the CVR *reading paths*.

The compositional system of *information value* expands the linguistic textual properties *given* and *new*. Kress and van Leeuwen (2006) developed a series of layout systems in which information is found to be organised semantically across texts. They analysed newspaper spreads and other layouts and found patterns of information where the left-side of the layouts approximated to that which is *given* (what we know) and on the right side *new* or "what the reader must pay particular attention to, of the 'message'" (Kress and van Leeuwen, 2006, p.180). Vertical relations also manage information value: in the *ideal-real* construct, the bottom portion relates the viewer to the *grounded* aspects of the layout or 'what is', whereas the top portion, usually populated with imagery, construed more charged information or the 'what might be', which in more commercial context equated to 'the promise of the product' (p.186). Finally, the centre-margin arrangement organises information around a central 'nucleus' of importance, where the more peripheral elements are in various ways extending the meanings derived from the nucleus ((Kress and van Leeuwen, 2006, p.194). Information value systems require considerable re-evaluation in the 360° image context, as they are based largely on fixed 'zones'.

CVR as an omnidirectional and rotation- based image naturally problematizes matters in this regard.

The third compositional category of visual grammar is that of *framing*. Framing is used to integrate information in visual text as being either *connected* (continuous) or *disconnected* (discontinuous), through the splitting and conjoining of visual elements. Empty space, lines, objects, and planes of any sort can function to achieve the effects of visual framing. CVR presents opportunities for reformulation of the kind of framing devised by Kress and van Leeuwen. Where framed image producers can use the rectangular bounds of the image (as well as lens effects) to create abstract planes, spaces and edges between and around entities, objects in the 360° space are always 'resolvable' as recognisable objects. A line between two people in a framed shot necessarily becomes a window (or door, wall for example.) once the spectator has rotated to take it in. Additionally, framing in CVR must be considered in a three-dimensional and immersive sense, on front, around and behind the spectator (who are themselves framed).

The last textual property included in the study is the text's *reading path*. Regardless of a text's modality, it will be read in some sequence, over time. This is reflected in visual grammar's information system components (given-new, ideal-real, centre-margin), which to be comprehensible as such must be read in some sequence. Unlike the written and spoken word, images have no guaranteed path, and instead are analysed for their 'most plausible reading path' (Kress and van Leeuwen, 2006, p.205). The image adds indeterminacy to the reading path when compared to language and this is compounded further in CVR as the image enfolds the rotating spectator. Indeed, a common concern in CVR production is that the rotating spectator may miss meaningful elements as they peruse other parts of the scene. This is a paradoxical problem of CVR where the 'three degrees of freedom' afforded to the spectator works against CVR as *cohesive text*. Additionally, not all parts of the 360-degree space surrounding the spectator are as likely to be viewed, where the rear areas are less likely to be accessed, rotationally (Tong, Lindeman and Regenbrecht, 2021).

CVR's problematic reading path has generated theories and practical methods for 'directing attention' centring largely on the use of POIs to guide the spectator around the scene (Brown et al., 2016b; Brillhart, 2018a). 'Directing attention' resonates strongly with both salience and reading paths, and as such, non-SFMDA concepts developed for VR are useful here to augment the methods. Brillhart (2018a) has conceptualised immersive POIs on the basis of probability. She

uses the editing technique of shot-matching[8] to conceptualise the edit as a bridge from one POI at the end of a shot to the POI at the onset of a subsequent shot. Shot matching of POIs are a useful way to ascertain the editing choices made in the production of texts. As CVR uses long duration shots, the spectator's view must also be guided around the shot through vector-based resources. These resources are determined partly by concepts such as kinetic cues, where the POI moves within a shot. Studies in attention cues in CVR such as Brown et. al. (2016b) suggest that multimodal attentional cues (movement with concurrent sound for example) to be particularly relevant in determining attention and thus CVR's reading path. *Social gaze* also takes on a novel textual functioning in CVR as a resource for directing the spectator's attention. Social gaze is cognitive phenomenon involving a natural sensitivity to where others are looking (Kendon, 1967; Frischen, 2007). It has been studied for its effects in VR (Rubo and Gamer, 2021), with mixed results, but with an acknowledgement that its study would be more suited to the kinds of texts studied here; "abandoning the repetition of homogeneous trials and instead immersing participants in social situations as semantically rich as many situations we encounter in real life" (ibid. p. 312).

3.7   Register

Halliday described register as "a conceptual framework for representing the social context as the semiotic environment in which people exchange meanings" (Halliday, 1978, p.110). Register is as such the measure of a particular kind of discourse situation ('the social context' and 'semiotic environment') with a genre-dependent range of semiotic resources available for 'exchanging meanings'. Register analysis is systematised through the variables *field*, *tenor* and *mode*, redounding with the ideational, interpersonal and textual metafunctions respectively, where "field concerns natural reality, tenor social reality, and mode semiotic reality" (Matthiessen 1989 in Poynton, 1991). I will outline these register variables and point to how they are useful in this study. As background, Matthiessen (2009; 2015; 2019), Martin and Rose (2007; 2008) and O'Halloran (2019) provide the big-picture modelling of register, and Eggins (2004) and Mechura (2005) were useful in their descriptions of the register variables.

Drawing from the ideational metafunction, field concerns the multimodal experiential content, "the social action: that which is 'going on', [with] meaning in the social system" (Halliday, 1978, p.142). 'Social action' here is two-fold, it is the *field-as-subject matter* and as *semiotic activity*. Subject matter are the aspects of reality that the text interacts with, and semiotic activity are the

---

[8] CVR has the facility to rotate the view in postproduction, aligning ('matching') elements between shots so they will appear in the same screen-zone when viewed.

communicative 'goings on', including *narrating*, *reporting*, *surveying*, *expounding* and other such semiotic activities. As such, field is used to empirically determine what the films are about, beyond their 'top-line' topical description. Ascertaining the texts' field, also provides an empirical grounding for later discussions regarding *genre* e.g., how *Congo VR* and *Mamie's Dream* report, narrate, and expound on their respective subject matters within the generic confines of nonfiction.

The *tenor* variable captures the interpersonal dimension of discourse. This was conceptualised by Poynton in her three dimensions of tenor: *power, social distance* and *affective involvement* (Poynton, 1985; 1991; see also Eggins, 2004). Power manifests in the authority and formality of a text, social distance as the familiarity of a text, and affective involvement through the construction of emotional contact between spectator and subject. The dimensions overlap and draw on similar resources such as MOOD and modality where speakers enact their position (e.g., subjective, objective, needing, giving) and stance (being certain, or not). Visual tenor resources combine similar resources of distance, gaze, body language, stillness, and activation, as well as the vertical positioning of participants around the camera-spectator.

The third register variable, MODE [9] draws from the textual metafunction in its concern for structure. Its *channel* and *bandwidth* variables are used to characterise texts as instances of a semiotic technology, as information channelled through their particular materiality (e.g., audio, video, one-way, two-way). In this regard, MODE adds systematic clarity to the sensory aspects of media, discussed in the previous chapter (section 2.2, *Decomposing CVR*), such as the synaesthetic aspects of immersive theatre, and the 'corporeality' afforded by the moving image. Transparency is also analysable through channel and bandwidth, as it is predicated on the sensory contact with media. MODE's *medium* variable is concerned with way language is employed within the material constraints and affordances of the form. For example, regardless of the medium in question, participants can nonetheless speak with monologic, or 'written' or dialogic 'spoken' MODE that will have a different constructing effect on the spectator.

The MODE variable *Experiential distance* captures the variability in role of language and the ways language interacts with action; from accompanying action to reflecting on action. The following examples are adapted from Eggins (2004, p.91).

- language being embedded in immediate action (using language in gameplay)

---

[9] To avoid confusion, I will capitalize MODE when using the word to denote the register variable and with lower-case when referring to the modes as they are used as the low-level analytical units (immersive (spatial, kinetic, gaze, sound), voice-over and music).

- language commenting on immediate action (sports commentary for example)
- language reporting on non-immediate action (reflection on actual events)
- language constituting action (reflecting on actual/fictional events)

MODE is of particular interest when characterising a new form such as CVR. It provides a holistic account of media texts, from their technological makeup through the semiotic choices made in the negotiation of meaning within the constraints of the form. The 360° image modality can be probed, adding semiotic description to the experience of, and textual uses of, immersion, presence, place illusion and other psychological characteristics of the form.

3.8 CVR's Analytical Units

3.8.1 Rank

In SFL, rank refers to the nested order of its lexical components (clause, group, word, morpheme (Halliday and Matthiessen, 2013)). O'Toole (1994) extended the concept of linguistic ranking order to imagery in his functional explications of visual artworks, ordered as follows:

- The work (such as the painting)
- An episode (a grouping of people/things),
- A figure (distinct person or thing)
- A member (a part of a figure)

His schema was later adopted by O'Halloran in her systematic functional analyses of film (2004). Her visual ranking order is, from largest to smallest:

- The film's plot
- The film's sequences
- A film scene (multiple shots)
- Mise-en-scéne (a single shot potential)
- The frame (the specific angle that instantiates the shot)

CVR's lack of a rectangular frame presents novel challenges for structuring the 'grammar' of its image. Framing allows for a shot to be 'about' things in a specific way: wide shots of an entire scene and groups of people; medium and close ups of individuals; extreme close-ups of their details, all provide a nested hierarchy of visual phenomena. CVR is ambiguous in this regard with resonances with both O'Toole's and O'Halloran's systems. The similarity with O'Toole's systematising of painting's rank is in the CVR spectator's potential access to the entire image simultaneously. With O'Toole's filmic rank it is the intersecting of frame and mise-en-scene. The difference in CVR is the framing enacted by the spectator. The concept is introduced here; in the next chapter I will propose a CVR-specific reformulation of O'Toole and O'Halloran's ranking systems (section 4.4.4, CVR's Ranking Order, Table 6, pg. 90), which will include necessary comparative details.

3.8.2 Modes

At the lowest level of SFMDA's stratified model is the *expression plane*, where one interfaces with the text. It is the expression plane that differentiates one type of text from another based on its modality (examples being *written*, *sonic*, *visual*). As a point of departure for studying the higher

levels of discourse, suitable modes must be determined as units of analysis. For reference, this study uses the following modes: *spatial-proxemic, kinetic, gaze, voice, diegetic sound,* and *music* lenses for examining CVR texts. I will address modes in a general sense here and discuss these particular modes in the next chapter (section 4.3.1, Transcription: Theoretical Considerations)

Early mode theories exemplified by O'Toole (1994) and Kress and van Leeuwen (2006) assumed that, where modes are found in texts, that each mode has a defined and separate contribution to play based on the 'mono-modal' semiotic system that it instantiates. In this 'polymodal' phase, modes were seen as distinct and discrete entities, often broken along sensory lines, with defined systems of meaning that could be summed together in intersemiotic analysis. Lemke pointed out that modes, once incorporated into multimedia texts, do bring modal characteristics, but are highly context-dependent for their meanings (Lemke, 2005). This has come to represent the basic consensus in mode theory as modes, while having a degree of specificity, are not considered fixed stable entities prior to their incorporation into multimodal texts, rather they are subjective and indeterminate. The indeterminate nature of mode is considered by Kress et al. (2001) as a function of the community in which the mode is found to operate.

> "The question of whether X is a mode or not is a question specific to a particular community. As laypersons we may regard visual image to be a mode, while a professional photographer will say that photography has rules and practices, elements and materiality quite distinct from that of painting, and that the two are distinct modes."

(Kress et al., 2001, p.43)

Tseng (2013) reframes the subjectivity of a mode as a "context-determined meaning making entity […] defined depending on analytical purpose" (p. 40). Boeriis (2008) frames this ontologically in his 'mode instantiation model'. Modes, in his model are only ever *immanent* things, as 'cloud[s] of semiotic nuclear particles' (p. 243) waiting to be instantiated in text. There is, for Boeriis, *pre- mode potential* (p. 242) but only in a very general sense. A particular situation of use (TV broadcast, news photograph for example will have a *mode register* (following register as described above)) and only through their 'mode-register potentials' will the systems of choice therein determine a mode (p. 244). Boeriis's model seems at first glance to be tautological but not when it is considered in the process of studying a novel form like CVR. For example, the kinds of meanings that can come from *gaze* are 'mode-potential' and dependent on CVR's textual register, described briefly as MODE above. Immersion formulates what gaze is in CVR where certain aspects of its meanings are made moot (looking off screen for example).

Given that mode is considered partly intrinsically, and part extrinsically determined, the question becomes *how much* meaning a mode can carry that is intrinsic, and how much is determined at the point of its involvement with the other modes in a text. For Lemke (2005), the modalities involved will be *specialized.*

> "Language, as a typologically oriented semiotic resource is unsurpassed as a tool for the formulation of difference and relationship, for the making of categorical distinctions. It is much poorer (though hardly bankrupt) in resources for formulating degree, quantity, gradation, continuous change […] the topological dimensions of meaning"

> (Lemke, 2005, p.87)

Bateman and Wildfeuer (2014) add nuance to the discussion of mode from a perceptual standpoint. Their reasoning lies in cognitive factors where for example we do not see and hear as two discrete sensory processes, rather we perceive in a synaesthetic, multi-sensory fashion. In this regard, mode is conceptually unstable from a cognitive perspective. They go on to say however that modes have inherent capacities to 'support different uses' (2014, p.182). They invoke Gibson's theory of affordances (as does Kress (2010)) as it caters for both the objective qualities of an object, and the subjective lens through which an object will be assessed for its potential uses (Gibson, 1979; van Leeuwen, 2005). For Bateman and Wildfeuer, this does not amount to an equivalence with polymodality. Rather, they accept that modes may have access to or 'reach into a particular material substrate' (pg. 182). 'Material substrate' here being akin to *visuality*, *aurality*, *tactility* substrates that can effectively be accessed in different degrees by different kinds of modes, but that are not claimed to be mode in themselves. To put this in context, it could be said that the in the selection and analysis of the specific modes in this study, I aim to access aural, visual, and also the 'immersive substrate' of the selected modes.

As is evident, discussing a 'mode' implies an acknowledgement of how modes interact in text. This is addressed in multimodality through various meta-theories and models of multimodal meaning. This is no mean feat, and the complexity and virtues of the task was articulated by Lemke (2005):

> "how we can mean more, mean new kinds of meanings never before meant and not otherwise mean-able, when this process occurs both within and across different semiotic modalities (in other words, language, visual representation, mathematics, etc.)"

(2005, p.93).

Here Lemke is pointing to the complex nature of intersemiosis where modes in-situ must be considered for their *multiplicative* effect on each other. A *functional cross-multiplicative* effect of multimedia texts for Lemke, presents new possibilities for meaning making, where modes continually augment each other. SFMDA theorists have proffered models and frameworks to manage this complexity and to bed multimodality in SF foundational theory. For example, SFL's stratified model forms the basis for Lim's *integrative multisemiotic metamodel* (2004, p.222) where he conceptualises a common 'space of integration' between linguistic and non-linguistic modes (he uses graphics and typography). O'Halloran developed a similar, stratified model of intersemiosis for printed texts (2008a).

As a relatively young field of inquiry, multimodality continues to develop and model meaning across modes. The consensus at present is that multimodality is reflexive: modes can have characteristics and a degree of semiotic potential, but this potential is only realised in the context of other modes. Tseng's remarking that modes are subjective and analytically determined is a critical aspect of multimodal inquiry into a novel form. Studying CVR requires that suitable modes are subjectively determined prior to analysis, for their *meaning potential*. My rationale for choosing this study's modes is laid out in the next chapter (section 4.3.1, *Transcription: Theoretical Considerations*) informed by the theories laid out here.

3.9   Cross-modal Metafunctional Organisation

The metafunctions' universality has made them a feature of many intersemiotic analyses (Royce, 1998; Cheong, 2004; Lemke, 2004; Baldry and Thibault, 2006; O'Halloran, 2008a; Royce, 2015). *Metafunctional organisation* across modes informs parts of this study, where modes are found to reinforce each other along functional lines. Examples include ideational functionality reinforced, where the same processes are both verbalised or written and represented visually. Interpersonal functionality becomes reinforced across modes where *questions* (in language) are reinforced by *demand*-based images. Textual organisation differs from the other two functions as it is inherently organisational in nature. Also, the textual function opens up into other cohesion-based systems that I discuss in detail in the next section. As such I will discuss the ideational and interpersonal in this section and defer the textual to the following section regarding cohesion.

3.9.1   Ideational Organisation

Of the three metafunctions, the ideational function, has received the most scholarly attention for its organising intersemiosis. In its experiential component, the transitivity system incorporates an 'open class' of lexical entities lending itself to multimodality, where *participants, processes* and *circumstances* can be realised readily across the image-text divide. The logical component of the ideational adds the aspect of *conjunction* to intersemiosis, where lexical entities can be related logically across modalities (cf. van Leeuwen, 1991; Martinec and Salway, 2005). Barthes' (1964) seminal work on image-text rhetoric undergirds much of the contemporary intersemiotic studies that are ideationally focused (van Leeuwen, 1991; Martinec and Salway, 2005; Royce, 2015) His concepts of *anchorage* and *relay* describe text and image relations of dependency and co-operation (Royce, 2015, p.720). Subsequent studies such as O'Halloran's printed advert analysis (2008a) and Martinec and Salway's studies on image-text relations (2005) typify the kind of scrutiny given to the ideational component of intersemiosis.

O'Halloran (2008b) provides a conceptual frame where image-text relations are considered as processes of *convergence* or *divergence* of meaning. Convergence implies a singular, unified transitive meaning, through 'contextualizing relations of parallelism' (cf. Liu and O'Halloran 2009). On the other hand, where meanings diverge across image-text, they are seen as dissonant, and recontextualize each other. Gill (in Unsworth, 2006) frames the relations of images and text as *ideational concurrence* (p. 60) where content in the image and text form relationships of either *exposition* or *instantiation*. Exposition occurs where transitive elements are repeated across image and text, resulting in a redundancy of information (nothing is added in either modality). Instantiation involves the image or text presenting an instance (a 'snapshot') from a more general kind of information in the other mode. Martinec and Salway (2005) reformulate these kinds of relations and tie them directly to SFL's tactic system of *dependency* (with the necessary dimension of status). Martinec and Salway posit a series of *equal*, *unequal*, and *complementary* status relations (Ibid. p. 340-431) where equal status points to relations of exposition (as does Gill), unequal and complementary statuses reflecting Barthes' anchorage and relay respectively. Their complementary status further opens possibilities for new and complex semantic arrangements. Cheong (2004) frames the complementary status of images and text as having a 'bidirectionality of meaning' that opens an 'interpretive space' which will, depending on the amount of re- contextualizing required, result in a greater 'semantic effervescence' (p. 176).

3.9.2   Interpersonal

Royce's intersemiotic complementarity schema (1998) explores congruences between visual

grammar's interpersonal markers and SFL's interpersonal semantics. As such, visual contact (as offer and demand) combines with MOOD with potential for intersemiotic 'reinforcement of address' (p. 69). Royce uses a combination of linguistic *statement* and visual *offer,* to illustrate a reinforced address. Unsworth (2006) uses Cheong's (2004) print advertisement analysis to illustrate a similar phenomenon. In it, a smiling woman is pictured with gaze directed outward towards the reader, accompanied with the text reading "I get the feeling that M1 wants me to enjoy value – and enjoy life. Everything they offer is brighter, nicer and more fun!" (2004, p.170). The reinforcement of address here realises a congruent reinforced address of *demand*, in the sense that the woman represented can be construed as engaging the reader directly in her gaze and perceived verbiage (as personalised 'I get…wants me').

Royce's schema doesn't involve *social distance*, an interpersonal resource that can be used to engender interpersonal reinforcement across modes. Matwick and Matwick (2014) discuss the effect of *synthetic personalization* in the context of cooking shows, where the producers of the show uses personal pronouns 'you know?' to reduce the 'aesthetic distance' (Adema in Matwick and Matwick, 2014). This effect of closing the interpersonal distance can also be found in a very distinct manner through the use of deixis. Deixis is inherently contextual involving verbally 'pointing' to things in the visual domain through deictic pronouns (e.g., 'that' thing over there). As such it can be used to close the perceived gap between speaker, spectator and visual object being referred to. As a part of the system of *cohesion*, deixis is mentioned here for purely for its interpersonal overtones and elaborated in the next section.

3.10 Cohesion and Conjunction

There is natural overlap between the textual metafunction and other text-forming systems such as cohesion. Royce's intersemiotic complementarity (1998) addressed the textual aspect through visual grammar's compositional variables of information value, salience and framing reinforced through their verbal counterparts. Other kinds of cross-modal organisation are found, such as *Inter-visual synonymy* and *reading paths* as they emerge across modes. Royce's schema is useful for its underlying premises but is limited in the scope of this study as he is observing Intersemiosis between modes on a printed layout, where image and text are part of a visual gestalt. Other means are required to ascertain how reading paths emerge in film and CVR's aural and visual modes, as they unfold across time as well as omnidirectional space. 'Inter-visual synonymy' also requires re-interpreting where relations emerge asynchronously and between aural and visual modes (e.g., synonymy between spoken element, realised 'after the fact' in the visuals). Here the system of *cohesion* is more apt in describing intersemiotic

complements as it takes synonymy as one of its lexically derived text-forming properties.

In SFL, cohesion determines whether a collection of words and sentences constitutes *text* or is simply a co-occurring string of words. Halliday and Hasan considered cohesion for the ways that words and clause-complexes 'hang together' as a recognisable units of meaning (Halliday and Hasan, 1976), such as in the well-worn example:

> Wash and core six cooking apples. Put them in a fireproof dish (Halliday and Hasan, 1976, p.2)

The text is cohesive as the reader can easily establish what 'them' refers to and as such can establish that the words are making meaning beyond the clause itself. Cohesion is a non-structural semantic phenomenon operating across multiple lexicogrammatic systems: grammar, lexis, and logico-semantic conjunction. This makes cohesion apt in non-linguistic contexts also, where a verbal element can be cohesive with a visual item. For this reason, cohesion is central to many multimodal treatises and analyses (O'Halloran, 2004; Royce, 2007; Liu and O'Halloran, 2009; Thomas in Ventola and Guijarro, 2009).

Where one linguistic element is interpreted by reference to another, they make a *cohesive tie* (Halliday and Hasan, 1976). Where ties are made, they can take the form of *reference, substitution,* and *ellipsis*. Reference involves ties such as pronominals (she, he, it), demonstratives (that) and more general deictic references such as 'the' (implying cultural knowledge e.g., '*the* economy') and 'this' implying' local access to something in the shared situation (e.g., '*this* stone in my hand…'). Substitution involves swapping 'so' or 'one' for the referenced element (e.g., "why must you do it? Because I said *so*"). Finally, ellipsis, or *substitution by zero* omits the reference item (e.g., 'Get a new house? Given the market, I wouldn't [get a new house]'). Grammar-based cohesion such as in the examples provided, is based on *closed* classes of words. Lexical ties on the other hand are based on open systems, where lexical elements are realisable in the visual domain, this enables cross-modal cohesion. Kinds of lexical ties are as follows (taken from Halliday and Matthiessen, 2013, p.644):

- o *Repetition* of words: bear - bear
- o *Synonymy* between words: sound – noise, with the negative antonymy included (sound – silence)
- o *Hyponymy*: taxonomic classification between superordinate and subordinate members of classes of things: tree (super) – oak (sub). Also includes co-hyponymic relations between elements of equal status: oak (co) – pine (co).

- o *Meronymy*: relations of parts to wholes: tree (super) – branch (sub) – leaf (sub-sub). Also involves co-meronyms: (extending the tree analogy) the relations of one leaf to another.
- o *Collocation*: where one word is typically found in relation to the other (smoke – fire).

Martin (1992) reformulated cohesion at the discourse semantic level, enabling cohesive ties above the clause and sentence level. For example, his system of *identification* is used to track entities across the text, through *presenting-presuming* and *generic-specific* subsystems. Where an entity is presented, it is established fully in its own right where no prior or extrinsic knowledge is required (e.g., '*a* dog'). Presumed entities on the other hand presume prior knowledge of some sort for identification to be fulfilled (e.g., '*the* dog'). Generic entities are presented as classes of things (dogs), where specific entities are singularized (e.g., '*this* dog') (Martin, 1992, p.102). These systems combine with that of *phoria* to allow for items to be tracked in texts. This manifests as *endophoria*, where ties are made within the text, *exophoric* where the ties are made between textual elements and extrinsic situational entities (as with deictic reference above), and *homophoric*, where the cohesive tie presumes cultural understanding (Eggins, 2004, p.34). The systems of presenting and presuming also combine with phoria to determine the direction of cohesive references. For example, a presumed entity that requires a previously presented entity for it to be decoded, is an *anaphoric* reference. Where the presumed, referencing item precedes its presented counterpart, it is a *cataphoric* reference (e.g., "In brief, the soon widely held assumption was this: *man could understand the universe*" (Halliday and Matthiessen, 2013, p.625 my emphasis))

Martin's discourse stratum and it's cohesive system of identification lends itself to the study of filmic texts, as implemented by Tseng (2013), Tseng, Bateman et. al (2012; 2021) and Bateman and Schmidt (2013). Bateman and Schmidt formulated cohesive dependencies in the discourse stratum of films in their system of *Broad Syntagmatic dependency Types* (BST). This was a move away from Metz's earlier *Grand Syntagmatique* where the shot was taken as a unit of analysis. Where Martin abstracted the content plane of language, freeing up analysts to work across larger swathes of written text, theories such as Bateman and Schmidt's did the same for film. Martin's identification system accommodates visual and verbal entities, being tracked through film. Tseng (2013) extends his *generic-specified* system (see Martin, 1992), where filmic participants are specified by various means. Her analysis of the film *Supersize Me* (Spurlock et. al., 2004) involves participants being formally identified, thus *visually specified* through graphic descriptors ('lower thirds'). Her analysis of Hitchcock's *The Birds* (1963) tracks participants phorically as multimodal chains of reference (see Figure 8, below). The presenting-presuming system is used to incorporate visual cues, such as repetition and part-whole visual ties (as cut-

in and medium shots). Baldry and Thibault (2006) discuss the phoric nature of *gaze* in their exposition of multimodality in film. Gaze is regarded as having the cohesive dimension of directing attention from the gazer to the object of their gaze (p. 167). Tseng, and Baldry and Thibault, each building on Martin's systems, provide a grounding for considering CVR's cohesion, where the immersive spatiotemporal aspect of the form can be tracked methodically, using identity chains (cf. Tseng, 2013 p.78).
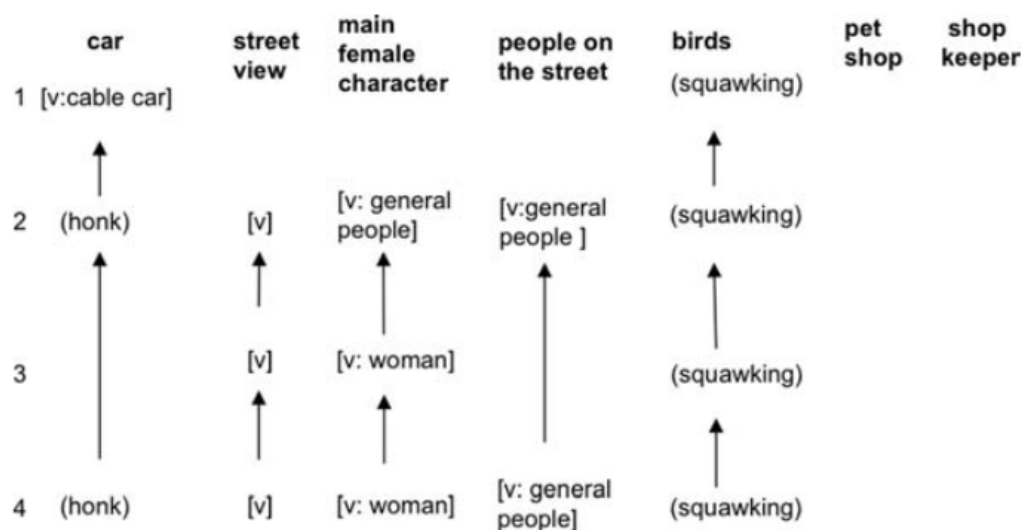
Figure 8: Example of cohesion illustrated as identity chains (Tseng, 2013)

Thus far cohesion has dealt with how items form referential ties. The other key aspect of cohesion is that of conjunction, and how text connects logically to build meanings at higher levels. The system of conjunction operates similarly from the clausal level upwards to inter-clausal formations and in the discourse semantic stratum. It is comprised of SFL's systems of *expansion* and *projection*. Expansion relates elements through *extension*, *elaboration* and *enhancement*, where extension *adds information* to a clause, elaboration *restates* (e.g., as examples and analogies) and enhancement *contextualizes* (Eggins, 2004, p.47). Projection, on the other hand construes thoughts (mental projection) and speech acts (locution). Halliday considered projection as incompatible on a rank level and did not include it in his system of cohesion (one cannot project something across sentences). This limitation is inherently linguistic and as such projection has been picked up and reinstated (albeit infrequently) in multimodal studies not confined to the sentence structure (cf. Martinec and Salway, 2005).

Conjunction was used by Van Leeuwen (1991; 2005) as a multimodal phenomenon in his analysis of documentaries. Extending Martin's *conjunctively relatable units*, he established conjunctive

cohesion amongst the verbal units, visual units and between verbal and visual units. In his analysis of *Industrial Britain* (Industrial Britain, 1931), he uses reticula, a graphical method favoured by Martin (1992) to visualise and track the ways the text expands its meanings across time and between modes (see example in Figure 9, below). This kind of analytical device is sufficiently flexible to be used in any media type, including CVR. Van Leeuwen's study is also important as he extends non-SF filmic cohesion theories such as Eisenstein and Timoshenko, as well as Barthes image rhetoric and Bill Nichol's contribution

to image-text relations.
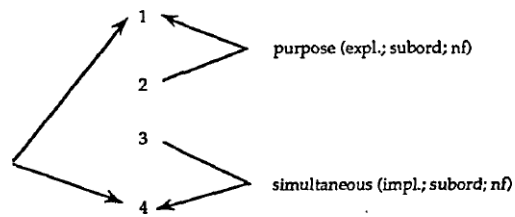
Figure 9: Example of conjunctive reticulum from van Leeuwen (1991)

3.11 Concluding Remarks

The goal of this chapter was to show the SFMDA methods 'from the ground up', to provide a sound foundation for the methods used in this study. Systemic functional frameworks have been described from their linguistic origins through to the current multimodal fields of enquiry, and multimodality has been discussed from its smallest units to larger systems of intersemiosis. As a social semiotic, the framework's suitability was described through the fundamental social functions that motivate language choices in the negotiation of discourse. The metafunctions were described as a means of analysing CVR, as its producers and consumers use its 'language' to address the same functions that initially spurred Halliday to develop SFL. I have described how SFMDA has adapted its resources from its linguistic origins, and outlined the semiotic systems already established across a variety of disciplines used as a conceptual through-line to analysing CVR. It would be impossible to include all possible approaches to this analysis and I have attempted to bring together those most apposite to CVR's novel spatiotemporal mix. As none of these semiotic resources are currently expanded for CVR analysis, they can at best be considered as 'close fits'. To fill the analytic gap, I have included some additional frames to extend extant semiotics into the CVR domain, such as Adam Kendon's theory of F-formations. This is a reasonable step to take in expanding visual semiotics and in keeping with the spirit of SFL, where "discourse analysis interfaces with the analysis of grammar and the analysis of social activity" (Martin and Rose, 2007). I have structured this chapter as a means of matching the framework with the methods, which are outlined the next chapter. This involves the metafunctional analysis, the register analysis, and the intersemiotic analysis respectively.

4. Research Methodology

This chapter outlines my research methods and how I will operationalise the social semiotic framework described in the previous chapter. I will begin with a brief description of the texts, followed with the rationale and methods used in their selection. Bevan and Green's *mediography of nonfiction VR* (2017; 2018) is described for its use as a starting point from which an objective filtering and purposive sampling approach was taken to arrive at the two texts, *Mamie's Dream* (Mamie's Dream, 2016) and *Congo VR: A troubled Past* (Congo VR: A Troubled Past, 2018). This process is informed by case design methodologies, which are discussed in the context of the choice of sample size and sampling criteria. I follow this by discussing analysis software in the context of studying immersive texts. I discuss the trade-offs in using certain tools over others and present my choice of analysis software, MAXQDA. On a related theoretical level, I address the question 'what is the data?' and state my rationale for considering the data for this study to be the material films as viewed in the headset.

CVR presents unique transcription challenges which I address from practical and theoretical perspectives. I describe the difficulty in deciding the amount of data to be transcribed, and the kinds of modalities best suited to transcribing CVR. I present the modes that are transcribed for analysis (spatial-proxemic, kinetic, gaze, diegetic sound, verbal, and music) and a brief rationale for their aptness in analysing CVR. The transcription's annotation and layout style are presented with examples. As a desktop study of publicly available texts, the study did not warrant a full ethical review, but I include a brief discussion in this chapter on researcher subjectivity and the ethical nuances of transcribing Sierra Leonean and DRC cultural phenomenon that I either cannot recognise or understand.

Having addressed the data gathering, I move on to the coding and analysis. The functional analysis is described as two-phased, addressing language-level discourse via the metafunctions, and addressing contextual discourse via the text's register. Coding practices are discussed, and the specific codes are enumerated and described.

I will preface the discussion by reiterating the research aims and questions addressed. The study is comprised of three overlapping aims: first, to describe CVR's immersive semiotics; second, to understand how CVR uses its resources in the production of nonfiction texts, and thirdly, to provide multimodal analysis with additional tools to analyse immersive texts. These aims are reflected in the underlying research questions.

1.   What kind of semiotic technology is CVR?

a) How does the CVR camera configure discourse in a manner different to framed imagery?

b) How does CVR discourse unfold across CVR's full modal ensemble*?*

2. *What are the implications of the CVR camera for producers of nonfiction texts?*

a) How can CVR nonfiction texts represent the social world?

b) How does the nonfiction CVR text construct the nonfiction spectator?

c) What are CVR's affordances and limitations for CVR documentarians?

3. *What augmentations to multimodal methods are required to analyse CVR texts?*

a) What additional analytical frames are required to bring current multimodal studies into immersive domains?

Broadly speaking, this chapter describes the methods used to address questions one and two. Question three differs in that it is addressed partly by the methods themselves.

4.1 Data Collection

4.1.1 The Texts

The texts chosen for the study are *Mamie's Dream* (Mamie's Dream, 2016) and *Congo VR: A Troubled Past* (*Congo VR*: A Troubled Past, 2018). *Mamie's Dream* is produced for Plan International (Plan International, 2022) and hosted on their UK division's YouTube channel (Anon., 2022). The organisation are, in their own words, "a development and humanitarian organisation that advances children's rights and equality for girls" (Plan International, 2022). The film centres on Mamie, a Sierra Leonean woman who narrates her autobiography, the arc of which culminates in her attending the 'Plan International's Learning Assistant's Programme' (LAP). The film has a promotional agenda, to raise awareness of Plan International's activities by showing how Mamie has benefited from their programme. Her story is also used to provide cultural and social context for the rights and roles of girls and women in Sierra Leone. The film includes a statement in the closing credits 'This is the true story of Mamie, spoken in her own words' where the film is positioning itself as a nonfiction artefact. The YouTube description includes language suggesting an experiential component "Take a walk in my shoes and experience what it's like to be a girl growing up in Sierra Leone" (Mamie's Dream, her full story: in virtual reality, 2016).

Congo VR: A Troubled Past (*Congo VR*: A Troubled Past, 2018) is the first episode in a series produced by the BBC's News division presenting the Democratic Republic of the Congo from several perspectives. The film positions itself as *news* both in its opening graphics and by its

inclusion in the BBC news YouTube channel (Congo VR: A Troubled Past - BBC News, 2019). This episode focuses on the country's history from colonial times to the more contemporary issues of self-rule. The production differs significantly from *Mamie's Dream* as it does not follow a single participant, favouring instead a general exposition of the country, in terms of its politics and physical and cultural resources. It is presenter-led and unlike *Mamie's Dream*, shows the mechanics of its production where the crew and their equipment are visible in many shots. There is a visible camera operator visible in many shots also suggesting that the VR content is produced in tandem with a conventional non-CVR production of the same subject. The film is hosted on the BBC News YouTube channel (Anon., 2022a). While it is under the banner of BBC news, the language used on the programme's YouTube page (Congo VR: A Troubled Past - BBC News, 2019) is similar to that on *Mamie's Dream*'s, locating it as a novel VR experience, "…Take a Virtual Reality journey through the Democratic Republic of the Congo from the Atlantic Ocean to the river's furthest reaches…" (ibid.).

4.1.2   Use of Case Study Design Frame

Prior to the selection of the texts, I conducted a pilot study on a segment from another CVR film *The Vodou Healer* (Walker, 2016) which is located in Appendix A: Pilot Analysis of The Vodou Healer. I needed to know what the implementation of a study such as this might look like, and to gauge the time and resources required to carry out such an investigation. Having completed the pilot study, I decided that focusing on two texts would be practical in the allotted timeframe while satisfying the primary objectives. To align this decision with the objectives of the study, I looked to case study theory. Given that the analytic methodology (SFMDA) is largely predetermined, case study concepts were useful as a means of rationalising the small sample size as well as the selection of the texts. Case study theory encompasses data gathering as well as analysis, in the case of the latter, theories such as Creswell (2014) advocate for a case-study-as-methodology approach involving things such as observations, interviews and reports. This study is more in line with Simons (2009) and others who frame case-study as the concern for *what is to be studied* (Thomas, 2011; Starman, 2013).

Where large-population studies are useful in testing theories through finding confirmatory patterns across data, case studies are more useful where the aim is to generate theory through the detailed examination of complex causality (Bennett George et al., 2005). Also, studies based on large sample sizes tend to isolate variables from their context to achieve stronger representative validity, for example by eliminating confounding variables. Where variables are unknown or undertheorized, such as in camera based CVR, it is more suitable to employ case

design methods that allow for things to be uncovered from the data, rather than deduced. As Becker (1990) states; "The entity or phenomenon under study emerges throughout the course of the study, and it is this surfacing that can bring the study to a natural conclusion" (p. 84). These sentiments reflect my own, as my study will not necessarily confirm anything but rather be formative in establishing systematic descriptions, as grounds for an emerging sub-field of inquiry. Essentially, case studies strive for depth over breadth. This has been described as an 'epistemological trade-off' that occurs between "the strength of a rich, in-depth explanatory narrative emerging from a very restricted number of cases and the capacity for generalization that a larger sample of a wider population can offer" (M Hammersley and R Gomm, 2000; and in Thomas, 2011)

Where a small sample size is conducive to rich explanations of phenomena, the question of which texts to study is also aided by case design concepts. Broadly speaking, the cases were chosen for their levels of overarching similarity and individual variability. They each represent a singular class of phenomenon: a camera-based nonfiction CVR production. The texts differ however in their contrasting styles used to represent their subjects, as reflected along theoretical lines such as transparency and reflexivity, which I will describe later in this chapter. As such, they fall along the dimensions of *typicality* and *diversity* as per Gerring and Cojocaru's case study typology (2008). To borrow from Eisenardt, they were chosen for their 'useful variation on the dimensions of theoretical interest' (Eisenhardt, 1989, p.296).

A point of clarification is required regarding the heuristic nature of the study. While the emergent approach taken is *heuristic* in principle and practice, that is a term reserved in case-study for the sequencing of case selection (Eckstein in Gomm, Hammersley and Foster, 2000) where the findings of one case's findings beget another case's selection, and so on - (Yin (2009) refers to this as 'replication logic'). While it is feasible that this study will suggest other cases to add depth to its findings, this would be outside of the scope of this study.

4.1.2.1   Case selection methods

To find suitable texts to reflect these 'dimensions of theoretical interest', I first required a universe of texts from which I could begin. Following that, I needed to isolate a smaller corpus of texts that exhibited the technical and textual features relevant to the question, namely those that allowed for a rich analysis of nonfiction CVR discourse, and that involves social actors in real social spaces. Bevan and Green's mediography of nonfiction VR (2017; 2018) provided the starting 'universe' of nonfiction VR. From here, a phased process of purposive sampling reduced the universe down to the two texts in question.

4.1.2.2 Filtering from the Mediography 'Universe'

Bevan and Greene's mediography was developed as part of the VR Documentary Encounters group (EPSRC, 2019) incorporating works created between 2012-2018. The mediography's database was developed through the researcher's first-hand searches as well as from other existing resources, such as the *MIT Docubase* (MIT, 2019) and the International Documentary Film Festival Amsterdam (IFDA) (Bevan and Green, 2017). As the first database of exclusively VR nonfiction works, the authors admittedly err on the side of caution with a 'broadly inclusive' approach to the titles (p. 163). As such, there is much technological and generic diversity, where CGI forms coexist with LIDAR scanning, photogrammetry, as well the filmic 360° video. Also, criteria for inclusion on the mediography included self-identification or identification by a third-party platform as 'nonfiction', resulting in a wide array of subgenres. While generating a slightly overwhelming array of choices, I didn't consider this problematic; the mediography's loose taxonomical boundaries simply reflect the highly porous field of nonfiction and the large volume of works, if anything, was reassuring in its comprehensiveness.

The mediography takes the form of a rich media website with a full catalogue index and search facility. It is essentially a listing of titles with a landing page for each one, useful in previewing titles to ascertain their suitability for this study. To begin, an Excel spreadsheet was created to capture the filtering and sampling process. Following brief communications with the authors of the mediography, I was given a way to select all films in the database and paste them into the spreadsheet from which point the process commenced.

At the time of accessing the site, there were 579 titles listed spanning the years 2012-2018. Each title has its own landing page containing a thumbnail image from the piece, a video clip as well as various metadata including the director and theme, festivals, and awards. The initial task was to eliminate non-CVR titles, generating an intermediate database of CVR works that were camera- based, and with minimal or no graphic components (lower thirds, inter-titles and similar graphical supports being common nonfiction elements, were allowed). Where further scrutiny was required, the embedded video clip on the mediography often contained the film in its entirety or a substantial part thereof as excerpts or trailers. Many titles however did not have any video clips available on the website and in these instances, an external search for the video was performed via Google, YouTube or on other proprietary VR platforms such as *Immersive* (New York Times, 2016) and *Within* (WITHIN, 2019). This filtering process was performed on a computer monitor as it would have been unnecessarily cumbersome to watch all 579 titles in a VR headset at this stage, where the criteria used could be ascertained quickly from a regular flat

monitor screen. This phase yielded 228 camera-based CVR films that formed the basis of the second phase of sampling.

4.1.2.3   Purposive Sampling Methods

The second phase of selection filtered the texts to reflect the research questions more directly, including only works of nonfiction that represented the social world. This resulted in the elimination of subgenres such as the nature texts *Arctic 360* (Arctic 360, 2016) and the *Planet Earth* trilogy (Planet Earth II: Deserts, 2016) which would not provide sufficient data for analysing the interactions of human participants. Spectacle-based and theatrical films such as *Dreams of 'O'* (Lajeunesse, 2017) and *Night Fall* (Night Fall, 2016) being purely performance-based were eliminated for similar reasons.[10] Following this phase, the spreadsheet contained sixty-two CVR works considered a suitable corpus of works for final sampling as they all reflected the core phenomenon to be studied.

Having arrived at a coherent corpus of works, a period of fieldwork and note taking commenced, where films were viewed multiple times in their entirety, in the HMD. Here, the spreadsheet was used to list the textual features in the films in relevant detail, approaching the core impetus of the study. This aspect of the sampling process was critical in finding texts with characteristics that could be interrogated for the configuring of discourse across modes as requirements of the first two research questions.

A set of criteria was established on which the texts were evaluated for their textual diversity along theoretical lines. A balance was sought in determining the criteria: too vague and they would make little meaningful differentiation between texts, too low-level and the criteria would begin to pre-empt the analysis resulting in under-cooked 'findings'. Using the field research, the research questions, and the nonfiction theories discussed in the literature review, texts were assigned the following rudimentary criteria.

- Narrator: where a narrated voice-over predominates the aural channel.
- Subject Speaks: the subject of the film addresses the camera.
- Inter-subject Dialogue: filmed participants engage with each other
- Talks to Camera: intentional engagement between filmed subject and the camera.
- Reflexivity and transparency (see below)

---

[10] I acknowledge that these are particular nonfiction genres in their own right and this stage of the process was question-driven purely to find texts most directly applicable to social semiotic analyses. I accept Nichol's classification of such films loosely under the *performative* mode (Nichols, 2010).

The last, global criterion deemed the text to be more, or less *reflexive* or *transparent*. This is relevant to studying the omnidirectional camera mode, where the crew must either consciously leave the site of filming or be a visible component of the shot. It also reflects the remediative aspect of emerging technologies (transparent immediacy and hypermediation) as well as the broad thrust of Nichols' documentary mode variations (Nichols, 2010), where documentaries consciously evince their own making, to varying degrees. Reflexivity is a broad term but, in this context, can be considered explicit where a presenter speaks directly to the audience, or more implicit such as with visible camera poles and/or production crew.

Having catalogued the texts and assigned each text with one or more of these characteristics, the films were chosen along the following lines. *Congo VR: A Troubled Past* (*Congo VR*: A Troubled Past, 2018) uses a mixed but largely reflexive approach, where the spectator is addressed directly at times by a presenter and by the filmed participants. *Mamie's Dream* (*Mamie's Dream*, 2017) takes a transparent approach with no acknowledgment of the camera by the participants. The film adopts a cinematic approach involving camera movement that extend classic tropes. The criteria, while blunt, served their purpose in finding texts that take a drastically different approach to production and have a wide range of voices, analysable in the texts. As shown in Table 1, below, the criteria lined up along the lines of narrator/subject speaks, the kinds of voices differ between the films. Where Mamie (as subject) is the sole voice-over narrating her own story in its entirety, the speaking subjects in *Congo VR* are (with one exception) visible and speaking directly to the camera. The most striking difference is perhaps the level of reflexivity evident in the films where *Congo VR* holds nothing back in term of the crew and their interactions with scenes, while *Mamie's Dream* allows no trace whatsoever of its production. This is summarised in the table below: the mediography sampling document can be found in Appendix B: Filtering Process from Mediography.

	Narrator	*Subject speaks*	*Int. Subj. Dialogue*	*Talks to camera*	*Reflexive/Trans*
Mamie's Dream	*	*	*		**Trans**
Congo VR	*	*		*	**Reflexive**

Table 1: Breakdown of textual features used in purposive sampling phase

As shown in Table 1, above, the texts chosen had sufficient variety in their styles and production modes. This allowed the research questions to be addressed in two significant ways. First, there is sufficient difference to get a comprehensive idea of the ways that CVR discourse can manifest in their different makeup. For example, how does a voice-over 'work' with a 360° image in

nonfiction representation, and how is this comparable to a diegetic direct address. Also, there is more scope for discussing the implications for the nonfiction producer when there is more variety in the textual resources being studied.

It should be clarified that even though both films are produced on the African continent (in the Democratic Republic of Congo and Sierra Leone), this was not an aspect of the sampling process. While the study will indirectly comment on how the texts and their producers intersect with African-based subject matter, this should not be taken as a foray into critical discourse analysis, of the representation of African themes. As I have described, the films were arrived at through filtering based on textual criteria. I briefly considered looking at a third film as a replacement for one of the texts but ultimately, I felt that to alter the selection process on the optics of their subject matter would be an artificial distinction to make and if anything, harmful to the integrity of the study.

4.2   Research tools used

There are no concrete examples of discourse analyses of CVR and as such, I had to develop a practical system of managing and analysing the data in the two films. The study uses MAXQDA as the main hub for analysis, together with Microsoft Word and Excel used for auxiliary language analyses. Other bespoke software was used for smaller aspects of the study (e.g., intonation analysis) and are discussed in their relevant subsections below. Approaching the study required a critical evaluation of the software tools available for a study such as this. Part of this involved identifying software used to manage similar studies to find the closest fit possible, but in practice many of the decisions were made as the project evolved as the affordances and limitations of the software on the evolving CVR analyses became apparent. As it is a novel study type, I will briefly lay out the decision-making process involved in their selection.

While some software products exist for annotating computer-generated VR (e.g., IRISVR (Anon., 2022c)), none exist at present that allow for CVR to be annotated and coded in the VR headset. It is within this overarching constraint that I decided to use a conventional qualitative analysis tool MAXQDA, as well as other auxiliary programmes to perform the analysis. The more fundamental theoretical CVR question in this light was what should be considered 'the data', and where it should reside. Ultimately, on a practical level, the question relates to how immersive data can be analysed in 'flat media' where distortion of the 360° image is unavoidable.

The pilot study (*The Vodou Healer* (Walker, 2016)) was performed in Microsoft Word and Excel

and having found this adequate for a small section of the film, it would not be sufficient for full text analyses. A such, I decided to use computer aided qualitative data analysis software (CAQDAS) in the main study, to handle larger amounts of data. This is perhaps not typical as CAQDAS are not necessitated by discourse analyses, where CAQDAS are considered by some to have developed with a bias towards grounded theory, which became popular at CAQDAS' inception (MacMillan, 2005; Paulus and Lester, 2016). Common CAQDAS such as Nvivo, MAXQDA and Atlas.ti, were tested as they allow for the integration of multimedia data. Multimodal analysis software TRANSANA and MULTIMODAL ANALYSIS were also trialled for their direct implementation of both multimedia data and built-in systemic functional tools. It was assumed from the study's inception that there would be challenges viewing, transcribing and analysing CVR texts as the frameless 360° image resists the box-like structure of software interfaces and their workflows. The trade-off it seems is that the multimodal programmes are supremely suited for video analysis, but at present this is limited to framed (non-CVR) video. The advantages of having inbuilt SF tools becomes moot when the video cannot be imported directly into the software. Attempts were made to circumvent this problem, such as converting the 360° image to a warped *equirectangular* version[11] to be subsequently imported into the software. This was quickly abandoned as it removed the interactive rotational component of the data while distorting the spatial relations of entities in the video. Another option was to extract portions of the image as stills and import those into the software, but in doing so decisions of what to crop out and extract would be pre-emptive of the analysis itself. Having accepted the reality that analysis software has not yet caught up with immersive video forms, I decided to view the data externally in its native display mode and to use MAXQDA to house the transcribed data for coding. The choice to use MAXQDA, a conventional CAQDAS, was based on its wide-ranging features which gave a degree of latitude for the analytical methods to evolve over the course of the study. Ultimately, a practice evolved of watching the films outside of MAXQDA in conjunction with the ongoing work in the software. The spatial aspects of the 360° image could be fully appreciated in the HMD and non- immersive aspects such as speech, fades and graphics could be viewed on a flat computer monitor. This was partly a matter of comfort and convenience which should not be underestimated when engaging in a study of VR texts, where the headset erases all else, including the analysts work-station. To smooth the process of watching the films in the HMD, I purchased a premium account on YouTube, which allows for the downloading and viewing of VR content offline, avoiding buffering and image quality issues.

---

[11] Mackenzie (2016) compares equirectangular warping to a globe being translated into a flat map.

From a theoretical standpoint, the decision to consider the HMD as the site of the data, is in line with video analysts such as Baldry and Thibault (2006) who consider the video itself as sacrosanct as the object of analysis. This is even more prescient in immersive texts, which both resist translation into other modes of representation, and lose their essence outside of the HMD. This sentiment echoes the historic consensus in discourse analysis that consider either the 'tape recordings' as the data, or even that the data was only ever to be found in "the naturally occurring speech in social life" (Zimmerman and West, and Button and Lee in O'Connell and Kowal, 1994).

The transcription process is discussed in detail in the next section, but I will briefly mention it here in parenthesis for its relevance to the selection of software. As a final field-test, NVIVO and MAXQDA were compared for their ability to render, manage and code large tables, a common manner of transcribing video texts (Hull and Nelson, 2005; Baldry and Thibault, 2006; Domingo, 2011; Bezemer, 2014). MAXQDA was chosen in this regard for its more advanced inbuilt word processing and flexible interface. In practice however, the amount of data gathered in the transcription process far exceeded any usable table structure in the CAQDAS and a different approach was later adopted, leveraging the internal document structure in MAXQDA to segment the films into a document-per-shot structure instead. Having completed the analysis, I consider MAXQDA to be a well built and well supported CAQDAS with a wealth of notetaking, analysis and visualisation features that became invaluable as the project developed. The specific uses of MAXQDA and the other auxiliary programmes are laid out in the following sections.

4.3   Transcribing CVR's Modes

Transcribing CVR texts for a multimodal study presented a particular set of challenges. The form distinguishes itself in highly detailed representations of heterogeneous social spaces. Crowd scenes enfold the spectator where voices, actions and interactions emerge from all directions. This, combined with the narrating voice-over and music soundtrack equates to the multimodal ensemble, all of which requires comprehensible transcription in a usable manner. The first challenge lays in establishing which modes are most apt in capturing CVR's heterogeneity, the second is in transcribing the right amount of detail to make sense of how the modes are functioning. All multimodal analysts face similar challenges in this regard and CVR perhaps amplifies some of the issues, where nothing is framed out in the conventional sense. With this in mind, I will discuss the methodological grounding and practical procedures used in the transcription of the films.

4.3.1   Transcription: Theoretical Considerations

The considerations for transcription involved what is transcribed, both in terms of the type of information and the amount of information, as well the procedure for capturing diverse multimodal data in a written transcript. There are practical, methodological, and ethical dimensions to consider, which are discussed throughout this subsection. I will begin by discussing the selection of modes for transcription. The intention here was two-fold: to cover a wide array of modalities and to pay particular attention to immersive-oriented modalities that would yield findings of interest to the specific way that CVR configures discourse in the 360° image. For this reason, I chose the following modes:

1. Spatial: proxemics, social space as well as physical space and its structures
2. Gaze: visual focus of participants towards and around the spectator
3. Kinesis: movement and actions
4. Diegetic sound: the live and atmospheric sounds recorded at the time of filming
5. Diegetic voices: voices captured in the location of filming
6. Voice-over
7. Music: the soundtrack and score

I chose modes 1-5 for their likely contribution to an understanding of CVR's immersive modality, predicated by the CVR camera. Each has the potential to operate in a way unique to CVR in that they are all spatial and can all 'do things differently' compared to their use in framed filming. For example, with few exceptions gaze cannot be 'off-screen' but must have a target, and space will involve the spectator's presence. For this reason, I will refer to them as the 'immersive mode group', as they are perceptually integrated with each other and co-occurring in the visual field. The other modes are more semiotically distant to these and were chosen to contextualize the immersive modalities within the filmic texts. By choosing these specific modes for the analysis, the mode of discourse that is uncovered in low-level analyses will be specific to CVR. As such, they address the first research question, directly.

Having identified the modes to be transcribed, the second concern was exactly what, in those modes should be transcribed. Many choices seemed self-evident where the text proffered clearly distinguishable characters performing clearly perceivable visual and verbal acts (Mamie in the boat; a Pygmy man digging, for example). Frequently however, the depth of transcription required a degree of subjectivity. All transcription is selective and based on the needs of the analysis (Davidson, 2009; Bezemer, 2014) but this was felt to a large degree in the more heterogeneous scenes (e.g., *Congo VR*'s city scenes and Mamie's church scene). Here, the sheer

volume of participants and their low-level actions resulted in choices of what to include, what to exclude and what to group together into larger units. In most instances the repeated viewings and iterative nature of the transcription resolved these questions. To catalogue the differences between units, I drew on aspects of multimodal theory such as visual rank (cf. O'Toole, 1994), and *clusters* (Baldry and Thibault, 2006).

Part of the challenge of transcribing CVR is missing relevant low-level moments in the data owing to their *naturalness* as multimodal events. Human cognition is very energy efficient and makes its gains by filtering extraneous low-level details while favouring 'best-guess' perceptions. Similarly, CVR's spatiotemporal mix can at times be overwhelming and to counter the sensory overload, I adopted techniques suggested by Bezemer and Jewitt (in Jewitt, 2016) such as 'denaturalizing' modes by viewing with the sound turned down and listening without the visuals. Repeated passes of the films yielded a set of transcripts which were further appended (albeit minimally) over the course of analysis where I found errors in the initial passes or where analysis suggested fresh angles and approaches to the modes.

On a theoretical level, the nature of transcribing visuals of any kind is an incongruent act, where images are *topological,* and their written descriptions are *typological*. Where images show continuous relations, words classify things (Lemke, 2005). There is a necessary process of transformation that has been compared to the semiotic effect of 'transduction' (Kress and Van Leeuwen, 2001). Lemke approaches this at a fundamental biological semiotic level where informational states continually alternate between typological and topological, as information is passed between systems. Topological states are continually translated into typological and vice versa: *"*Smooth motor behaviour (topo) organized as visual and verbal signs (typo): gestures, words, in ecosocial supersystem as metasystem of interpretation" (Lemke, 2006, p.108). Where the reverse occurs, he uses the following neurological example*: "*Firings in neural nets (typo) organized as coherent cortical effects (topo): brainwaves, holograms" (ibid.)*.* While he is not referring to image and text translation per se, Lemke's theory is interesting in his assertion that for information to be interpreted, it *must* be translated from typological to topological or vice versa "In a mapping of continuous variation onto continuous variation, there is very little room for novelty or innovation; there is only redescription" (ibid. p.107).

Finally, ethical considerations were considered regarding researcher subjectivity. Transcription is always somewhat political as it is partly a process of encoding the subjectivity and biases of the analyst (Jaffe, 2000). Transcribing CVR is unique as the immersed researcher (as spectator) engages in a process akin to transcribing one's own surroundings, being made to

feel present into an unfamiliar territory. The chosen texts' abundance of visual detail contains cultural details also, some of which will inevitably go unnoticed. As a male, European, spectator-researcher who has not visited either of the texts' locations, I can but encode what I see through my own cultural lens. As I have alluded to earlier, this study is focused on the mechanics of discourse and representation than on the cultural specifics of Sierra Leone or the DRC. I am certain to miss some aspects of the scene, but this is not a significant worry in this study. Another way of considering this is that the audience for the texts is by and large intended to be an uninitiated western one, and a member of that audience, what I 'see' in the data will reflect the intended entry level for the films.

4.3.2   Transcription Process

As discussed above, the films' transcripts were housed in MAXQDA for coding and analysis. Each film is represented as a folder, containing text files, where each of the films' shots are represented as a single file (see Appendix C: Examples of Transcribed Shots). MAXQDA has a panel-based interface layout, in which the 'document system' panel acts partly as a navigation system, allowing for quick access to the shots (see Appendix C.3 MAXQDA Interface). In this section, I will describe the procedure of getting the data from YouTube, into the transcription documents. I will describe the final transcripts in terms of layout and annotation methods used. The multifaceted nature of the process required some diverging paths of transcription and annotation before being housed in the full final transcription. This was partly due to the way language is parsed in SFL analysis, which has its own favoured practices (cf. Eggins, 2004) more applicable in grid-based programmes such as Excel. This complicated the linearity of the process, resulting in some phase-shifts in terms of where annotation ended, and coding began.

As mentioned, YouTube was the source of the data as it hosts both films, each available at 4K resolution. During transcription, I alternated between the browser and an Oculus Gear VR headset, where necessary. Transcribing the sound was done directly on the YouTube website which presented no information loss as both films use a stereo mix that sounds the same regardless of whether the HMD is used. Transcribing the immersive modes however (space, kinesis and gaze) required the HMD as viewing CVR on a flat monitor warps the footage into the equivalent of a very wide-angled ('fish-eye') lens, the optical effect being that perceived distances become exaggerated, making transcription of the space mode inaccurate.

Two tracks emerged in the process: I transcribed and annotated the visuals and non-verbal sounds directly from YouTube; the linguistic modes required an intermediate SFL annotation

phase. The annotation of the immersive modes were adopted mostly from Baldry and Thibault's (2006) transcription of video texts, with necessary modifications made to reflect the needs of the 360° environment. The goal at this stage of the process was to keep a consistency with established annotations and to keep the system clean and intuitive. Throughout the entire transcription process, Ehrlich's criteria provided guiding principles: simplicity and validity, good readability and correctability with a minimum of transcriber and user training (in O'Connell and Kowal, 1994, p.84). Below, is an example of a transcription of the visual immersive modes, comprising an extract from *Mamie's Dream* shot #3. This includes the annotated modes: kinesis, spatial and gaze. The transcribed segment shows the overall layout, how the modes and their transcribed entities are segmented and how they are represented temporally. Data is shown as a cascading series of nested brackets denoting clustered, consecutive information. For the full annotation key, see Appendix D: Annotation Key. I will take each of the transcribed modes in turn, beginning with *gaze*, and provide more clarity on how they are annotated.

Gaze is annotated as either *engaged* or *disengaged*: where engaged, the object of the gaze denotes the gaze vector (e.g., CROWD: ^ watching FOOTBALLERS). Where gaze is disengaged, as when one is internally focused or 'lost in thought', the object is 'undefined' and the direction of gaze is used instead ('upwards', downwards'). Where a gaze is partly a reaction to an external stimulus, the carat '^' symbol is used to link stimulus to reactor.

Gaze:
FOOTBALLERS: ^ at ball and each other (engaged) CROWD: ^ watching FOOTBALLERS

Kinesis is annotated in a heuristic manner, as a description of transactional processes (e.g., FOOTBALLERS Kicking and following ball from distance across camera position (below)). Movements occur sequentially and concurrently, and with many overlaps. Sequential movements were segmented using semi-colons, with concurrent movements grouped within square brackets. Where a series of movements are involved as action-reaction sequences, they are grouped in round brackets and separated with caret signs. The temporal nature of the movement is further characterized by as its tempo (s: slow; m: medium; f: fast).

Kinesis:
[
FOOTBALLERS Kicking and following ball from distance across camera position (f)

BALL rolls in zig-zag directions between passing players ^ PLAYER #1 runs to meet ball; kicks ball to side-line ((behind rocks)) (f)
]

[
(PLAYER #1 runs to where ball has settled; kicks ball to PLAYER #2) ^ (PLAYER #2 kicks ball
along side-line) ^ (PLAYER #3 kicks ball into distance (f) ^
        (FOOTBALLERS walk and jog towards general area of play (s>f)
]

CROWD ((mixed)) standing (mostly), running, walking

The *spatial* transcription corresponds to the embodied camera, its horizontal and vertical position, the distances between it and filmed participants, as well as between participants. As with kinesis, space was described partly through heuristic description (as seen in [CAM P] in the example below). Camera height is broken into five potential values, from extremely low to extremely high, encompassing heights from ground level to aerial perspectives. Relative distances between entities are annotated with connecting angle brackets '<>' (as in the [CAM <> FOOTBALLERS] example above). Distance values were derived from Hall's proxemics (intimate, personal, social, public). To fully utilize the spatial properties as per visual grammar, as distance and involvement, the orientation of the participants was included, as 'front', 'side' and 'away' with oblique angles denoted as 'front-side' and so on. Finally, regarding the spatial mode, an 'activated' variable describes the horizontal zone of active space around the spectator. This will vary from the placement of the camera-spectator either in a 'fully activated' space (e.g., [Activated] 360°, below) with all directions containing action or the potential for action, or a more enclosed space where the action is restricted to a narrow portion of the spectator's field-of-view (e.g., Mamie #7 [activated vert.] enclosed: low ceiling with material hanging).

Spatial:

[CAM P] on ground, rocky patch between edge of football pitch and school-grounds elevation [CAM H] medium (to football pitch); sub (to school-grounds)

[activated] 360-open

[CAM <> FOOTBALLERS] public/social/personal; mixed [CAM <> CROWD] public; mixed (mostly front)

Following Baldry and Thibault's video analysis, a gloss of the visuals was added to the top of each transcribed shot, giving a general description of the shot's visual content. Again, this follows and was used for additional information tangential to the study such as graphics as well as the visual transitions. The extract below contains the visual gloss of the same shot in the example above. The gloss was also useful in general notetaking.

[Cut]

Mamie not present. Boys playing football on clay-like surfaced football pitch. Small incline separates playing area from raised area with buildings and telecommunications mast. Small crowd lines the top of the incline looking down at the footballing boys

[\dissolve: s]

The sound was transcribed differently depending on the mode, voice-over, diegetic sound, or music. The diegetic sound follows the same notations as used in Baldry and Thibault's schema but as a discrete transcription component, the non-verbal sound did not require annotation differentiating it from other sonic modes. As such, the principal notation required as shown below is the *origin* of the sound, the volume it is heard and whether it carries across from one shot to another. The music was annotated in a similar fashion, as a description of the musical field, its sequencing, and its volume in the overall mix.

Diegetic Sound:

lively crowd-sounds (vol:m); rhythmic knocking sound (vol:l)(cont...)

Music:

rhythmic knocking continues and is extended by intro of salient drum beat
rhythmic percussive elements; simple piano melody single 'note'; harmonic notes added
(vol:m)

The voice-over and diegetic speech constitute the *linguistic modes* of the text and can be grouped for the purpose of description here, both being annotated using SFL. To begin the process, the verbal components were captured from the films using a combination of YouTube's auto-captioning facility as well as through manual transcription in Microsoft Word. Having captured the wording, I was faced with the decision of whether to code the linguistic modes in MAXQDA using SFL, or to perform a preliminary analysis in an external application. I opted for the latter approach for practical purposes, as linguistic analyses seemed more visually appropriate to horizontal and linear formats such as Excel. As such, I annotated the linguistic along its SF dimensions in a series of Excel spreadsheets prior to importing the results into the analysis proper in MAXQDA. UAM Corpus Tool was used to partly automate this process, in conjunction with Microsoft Excel, where additional manual analysis of the metafunctions was performed based on Eggins' exposition of SFL (2004) and Halliday and Matthiessen's functional grammar (2013).

The lexicogrammar and intonation were represented in the transcript in the following fashion, listed by functional system and property. This aspect favoured a labelling approach and as such the clauses were broken into their respective metafunctions and provided in a labelled list

format, as in Table 2, below. The extract used in this example is taken from C#4, "I'm Alastair Leithead the BBC's Africa correspondent. And the team and I will be taking you on a great adventure up this magnificent River. We'll be exploring its history, its riches, its poverty, and its future"

Ideational	Interpersonal	Textual
Participants: 1) Token: I; identity: Alastair Leithead ('I'): BBC Africa Correspondent. 2) Actor: The team and I Recipient: you (spectator) range: a great adventure 3) Actor: We Scope: this magnificent river; Range: its history /riches/ poverty/ Future **Processes:** Relational identity: (I) Material: 'taking you' Material: 'exploring' **Circumstance:** Location 'up this magnificent river'	I'm (AL) The team and I will (be taking you) We'll (be exploring)	(top: I) > 'm Alastair Leithead) the BBC's… (txt: and)(top: The team and I) > Will be (top: We) > 'll be exploring…

Table 2: The presentation of lexical metafunctional data in the transcript

The language aspect of transcription was perhaps less straightforward than the visual modes as I required the voices to be probed through several functional systems, resulting in multiple transcriptions (see examples below). For example, the verbal track required analysis of its metafunctional variation both in the lexicogrammar, and in the intonation (see examples 1 and 2). I also needed to have a usable transcript for analysing the conjunctive relations between clause complexes (example 3). For this reason, there are multiple versions of the spoken word in the transcript, each annotated along different dimensions, as illustrated below. This was necessary but cumbersome as there were multiple representations of data making coding difficult to manage.

1 - Prosody (using the non SFL Jefferson system)

>I'm Alastair Leithead the BBC's Africa correspondent< (.5) and the <u>team</u> and I will be taking you on a great↑ adventure ↓ (.) up this magnificent River (.) we'll be exploring its <u>his</u>tory (.) its <u>ric</u>hes (.) its poverty (.) and its future
↓

2 - Intonation (SFL)

.1. I'm /Alastair Leithead /the BBC's Africa corres/**pond**ent // .1+ And the /**tea**m and I // (4 contingent) will be

/**tak**ing /you // 1+ on a /**great** /adventure // .1. **up** /this /magnificent /**Ri**ver // .1+ we'll be /exploring its /**hist**ory

// .1+ its /**rich**es // .1. its /**po**verty // .1. and its /**fut**ure

3 - Logico-semantic

(a) I'm Alastair Leithead // (x b: feature) [who is] the BBC's Africa correspondent

+ || and the team and I will be taking you on a great adventure up this magnificent River

+ || (a) we'll be exploring its history // (+b1:expand-add) [and exploring] its riches // (+b2) [and

exploring] its poverty // (+b3) and [exploring] its future

The Logico-semantic and intonation dimensions are inherently sequential in nature and required linear representation in the transcript. These intermediate analyses were performed in Microsoft Word, purely for ease of word processing, and then pasted into MAXQDA. The SFL system of intonation was annotated for pitch patterns and tonal groupings, aided by WASP speech analysis software (Anon., 2022b). Example two above shows an intonation-based excerpt from *Congo VR*. The slashes segment the speech into tonal groups and the numbers denote specific pitch patterns. These numbers were used to annotate whether the dominant (tonic) syllable is inflected downwards ('1') upwards ('2'), level ('3') or a combination of these, such as in '5' which is a rise- fall pitch pattern for example, indicative of *committed* statements (O'Grady, 2017).

For the logico-semantic dimension (example 3), the mode of expansion and projection of clauses was denoted using conventional SFL notation. Letters in parenthesis denote tactic relations, identifying whether a clause is for example, dependent ('a') or independent ('b'). For expedience, I opted to use the Latin alphabet over the more cumbersome Greek alphabet as is typical in SFL. Other annotated elements include the '+' which signifies relations of extension, 'x' as elaboration and '=' as relations of enhancement (Halliday and Matthiessen, 2013, p.447).

4.4 Coding and Analysis

Having selected and transcribed the texts' modes, I began a functional analysis undertaken in two broad phases: as *language* and *register*, reflecting the stratified nature of SFMDA. The language phase focused primarily on the metafunctions, where CVR realises low-level meaning in its immersive modalities. The texts were then considered in this phase for the broader processes of Intersemiosis *across* modes, using metafunctional organisation principles and theories of

cohesion and conjunction. The second, register phase builds a picture of CVR as situated discourse, bringing the analysis into a contextual domain where the study could interface with nonfiction genre.

Both phases of the analysis were performed primarily as a coding process in MAXQDA, and where required, moved to a Microsoft application. As the data-proper was at all times considered to be in the HMD, analysis was performed as a series of moves between the HMD and the software, as immersive viewing, and coding. I will outline the processes used in both phases of analysis from initial coding to the reporting of my findings. I will discuss the theoretical grounding for the coding practices used as well as the other auxiliary practices adopted in analysis.

4.4.1 Analysing CVR in the language stratum

The aim of the analysis in the first phase was to ascertain the kinds of low-level meanings that would manifest through the metafunctions, addressing the first research question's concern for the configuring of discourse in CVR. Implicit in this is the effect of the camera on visual meaning, and the way the novel 360° image is used in the wider ensemble of modes to create larger contextual meanings. From a technical standpoint, the phase also elucidates how the semiotic resources used in analysis of framed texts, morph in CVR, thus addressing the third methods-based question. For example, how a visual process is encoded differently as a functional system when it occurs across and around the spectator's position in 360° space.

Where coding theory is concerned, comparisons can be made with existing coding methods, as catalogued in Saldana (2012)). At a high level, the study uses a *hybrid* coding approach that is both deductive and inductive as it applies (deductively) and extends (inductively) SFMDA systems throughout the analysis. Azungah (2018) describes the deductive approach as beginning with 'an organising framework comprising of themes for the coding process' (p. 391). This study's 'organising framework' is the set of SFMDA resources used to build the initial codebook. In this sense, the process is comparative with *structural* coding practices, as I am both coding and categorising the data with a priori concepts (Saldana, 2012). 'Structural' categorisations, however as laid out by Saldana involve more arbitrary question-based systems, rather than methodological frameworks, such as SFMDA. The inductive aspect of the study resonates with Thomas (2006) who describes inductive coding as "approaches that primarily use detailed readings of raw data to derive concepts and themes" (p. 239). The inductive aspect of the study is largely relating to the third research question, where novel configurations of discourse in the raw 360° data, retroactively reflects in the updated codes. I consider the analysis used in this study as a dialogue of sorts between the coding system and the data, where

deductive and inductive methods reflexively adapt throughout. In this regard, the coding is itself an analytical process, a (non)distinction made by Miles and Huberman (1994, p.56) who consider coding and analysis to be a singular process. My aim on reaching coding saturation was to have a set of codes that describe a tentative functional system for CVR texts that could be drawn out in the findings for its significant modifications of the extant semiotics of film, imagery, and other forms.

Following a phase of fieldwork, which also benefitted from the pilot study, an initial codebook was developed based on extant semiotic systems from domains considered potentially useful to analysing CVR. The full code book is available in Appendix E: Code Book and Trees. Many of the codes related to well established SFL resources used in the linguistic analyses, and as such came as a pre-existing 'kit' of resources. The nonverbal semiotic resources used as codes, were adopted indirectly from related visual forms. Kress and van Leeuwen's *visual grammar* (2006) provided many of the visual codes (*offer, demand, visual process types* for example) and was foundational to the visual modal analysis. Other less obvious systems were adopted where I felt they would contribute to the *immersive* modality engendered by the CVR camera. Architecture (Stenglin, 2009), for example seemed a natural fit, in that it conceptualises its semiotics around an imagined 'reader' that is immersed in physical spaces.

Defining at the outset, what constituted a 'semiotic resource' presented certain taxonomic challenges when setting up the codebook. This is partly due to the many different perspectives I drew from in the study. Also, because SFMDA is a metatheory used in multiple forms it does not have a singular system or set of terms to pull from, like SFL does. An example of this is the use of the code *perspective*, which is an essential descriptive tenet of visual grammar but not a technical term used in SFL (although it is connoted through other means). Initial attempts to resolve these discrepancies ran aground, and rather than continue the impossible task of resolving verbal and visual grammars into one system (essentially 'solving multimodality'), I allowed for codes to be created based on their semantic usefulness. As such, I included low-level resources as found in SFL and in multimodal frameworks, and concepts that were related to these resources. Perspective was thus included as 'meta-resource' and a valuable descriptive category which also works as an organizing feature under which other functionally apt sub-codes could be located, e.g., vertical positioning, movement, objective-subjective arrangements in the visual as well as linguistic resources such as deixis. Also, where issues like these arose, the code system was allowed to flex throughout the analysis, where problematic categories were modified.

81

The code system, as it pertains to the language-level analysis was broken into three top-level codes equating to the metafunctions using both linguistic and multimodal terms:

- Representational/Ideational
- Interpersonal/interactive
- Compositional/Textual

Each of these top-level functions contained a nested hierarchy of corresponding semiotic systems: *Interpersonal/Interactive > Contact > Offer*, being an example of a 'code-branch'. Figure 10, below provides an example of the interpersonal coding system, as it derives from systemic functional resources.
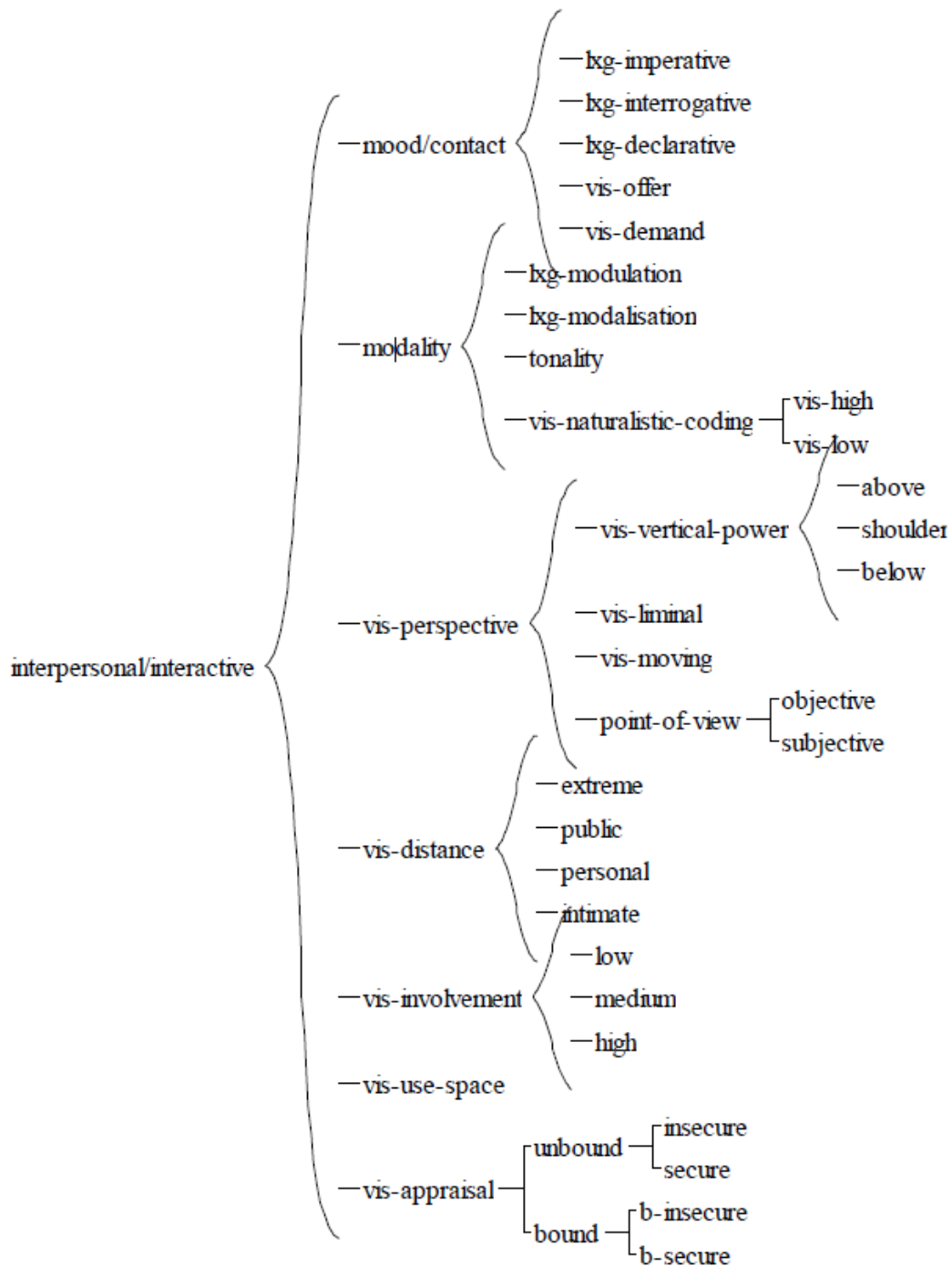
Figure 10: Example of code tree drawing from SF interpersonal resources

Having identified the code system, the ensuing analysis involved parallel processes of applying codes and memos to the data. The codes allowed the modes to be categorised along functional lines, whereas the memos were used to make exploratory free-text, observations on aspects of the data. Memos allowed for sections of the data to be highlighted and noted for their novelty and were formative in making the adaptations in the code system. Memos were applied at multiple

levels in the software: in the transcript; directly on individual codes in the code tree; and on the shot documents that housed the transcripts. The code-memos also contained a synopsis for each related semiotic resources, allowing quick reference and updating as I found new perspectives on respective theories. Finally, the logbook feature was considered the highest level of memo, a text editor reserved for moments of discovery that would have global consequences in the study.

4.4.2   Intersemiotic Analysis

Intersemiotic analysis was broken into two related phases, a *cross-modal phase* where shots were analysed for meanings that emerged between image and voice, and a *cohesion phase* where cross-modal analyses were extended across macro phases of text to identify how CVR's discourse is produced in a cumulative fashion.

Cross-modal analyses, being local to singular shots could be performed using the same overarching process of coding and applying memos to the transcripts. Coding was applied for the low-level relationships between image and voice, using relevant intersemiotic theories to formulate codes (Martinec and Salway, 2005; Royce, 2007; Liu and O'Halloran, 2009). Deriving codes for this part of the analysis required a looser approach as intersemiotic meaning is multi-systemic, incorporating the three metafunctions as well as the system of cohesion. With this in mind I used the following top-level codes:

- Ideational/cohesive
- Interpersonal complementarity
- Logical conjunction
- Cross-modal dependencies

The last code 'cross-modal dependencies' involves concepts derived from Barthe's anchorage and relay (Barthes, 1964) which is subsumed into Martinec and Salway's *dependency systems* (Martinec and Salway, 2005).

Where the macro analyses were concerned, the dual systems of cohesion and conjunction were used to track meanings across modes and shots. For the cohesion analysis, a discourse phase from each text was selected and analysed for the identity chains that track cohesive entities between shots. Tseng's filmic cohesion analytical methods (2013) were applied to each phase, providing a graphical representation of the cohesive ties binding the phases together. This portion of the study required an open, grid-like workspace to represent the films temporally and as multiple

84

co- occurring chains of information. A such, I moved intermediary portions of the analysis to Excel. Where conjunction analysis was concerned, the same approach was taken. Taking van Leeuwen's conjunctive studies of film (1991) as the theoretical basis, the texts were analysed between MAXQDA and Microsoft Word, where a tabular approach was required. Both the cohesion and conjunction analyses were represented graphically: Tseng's filmic identity chains were used for the cohesion (see Figure 8, p.59 for example from Tseng's analysis (2013)), and van Leeuwen's reticula were used for conjunction (see p.60) In both cases, the visualisation of the systems provided a fresh insight into the findings, leading to an iterative process of discovery.

4.4.3   Analysis in the context stratum

Register occupies a more abstract domain than the metafunctions, and analysis of register was multifaceted and less linear than the language analysis. The analysis began in MAXQDA where the transcript was coded with register-specific variables, such as the top-level register variables of *field*, *tenor*, and *mode*, as well as lower-level properties such as *lexical density*. Much of register is concerned with how its variables realise higher-level contextual features of genre, and as such the work in MAXQDA was augmented with auxiliary analyses done in both Microsoft Word and Excel.

The *Field* variable was analysed along its two dimensions: as *subject matter* and the *socio-semiotic activity* used to manifest the subject matter. As the field variable is concerned with content, it drew from the coding in the ideational metafunction. Patterns and frequency of participants provided the subject matter while the semiotic activity was realised through processes and circumstance types. An example of subject matter realisation is where 'father', 'mother', 'families here in Sierra Leone' and 'stepmother' all realise a subject matter of *family*. The process of determining subject matter involved finding similarities in participants, both human and nonhuman, and sorting them into higher-level phenomena. Visual participants were included in the taxonomy with 'v' in parenthesis to denote *visual* field. As the process of determining subject matter was largely of identifying, listing, sorting, and counting entities, I combined the ideational coding in MAXQDA with Microsoft Excel for this part of the analysis. Once all participants were included in the spreadsheet, I reduced the large and heterogeneous array of participants to a manageable taxonomy of subject matters. Each subject matter was considered complete when it had sufficient weight and frequency of representation in the overall text. A final taxonomy of subject matters was ordered by its size, roughly equating to a 'primary', 'secondary' subject matter and so on down. A full breakdown of the texts' subject matter is found in Appendix F.2:

Subject Matter: Congo VR and Appendix F.3: Subject Matter: Mamie's Dream. Socio-semiotic activity was derived from the processes and circumstances. *Reporting*, for example is implicit in material process types such as in 'they blindfold you...', and 'I sat six subjects'. Circumstance types contribute to both the socio-semiotic field and to genre, where for example circumstance of *time* is indicative of generic staging of *narration*. Matthiessen's field typology was used as the basis of this part of the analysis (Matthiessen, 2015). His typology covers many text types including those not relevant to this study (shopping catalogues for example) and his examples required slight modifications to be used in film analyses. Through a preliminary process of testing the typology against the two films, an adapted set of socio-semiotic activities were derived based on Matthiessen's typology, maintaining their essential characteristics. The socio- semiotic activities that ultimately formed the analysis were *reporting, expounding, recreating, exploring, advising, sharing,* and *doing*. An example is provided here of *reporting* including the modification made. The full list of socio-semiotic activities is in Appendix F.1: Socio-semiotic Field Types including their modifications from Matthiessen's initial taxonomy.

Reporting	
Matthiessen's original definition of socio-semiotic activity (2015, p.8)	As used in this study
"...reporting on our experience of particular phenomena — chronicling the flow of particular events (as in historical recounts or news reports), surveying particular places (as in guidebooks) or inventorying particular entities (as in catalogues)"	**Chronicling**: reporting <u>temporal</u> events as part of the flow of observable history **Surveying**: to report/examine/describe (<u>nontemporal</u>) entities for their features. This is extended to people, places, and things. *Inventorying not included*

Table 3: Example of socio-semiotic activity (reporting) adapted from Matthiessen (2015)

Tenor: Interpersonal resources were used as the basis of analysing tenor where the enactment of visual, verbal and intersemiotic relationships were coded in the language stratum. MOOD and modality in the verbal mode, as well as contact, distance and perspective in the visual modes combined contributed to tenor. Each of these resources were framed against Poynton's three dimensions of tenor (Poynton, 1985; see also Eggins, 2004).

- Power: the status relations between the spectator and the text
- Contact: the nature of the role relations between spectator and text
- Affective involvement: the emotional expectations placed on the spectator

The three dimensions overlap and coalesce into other tenor features such as *familiarity* where a tenor can manifest as *equal status, frequent contact*, and with distinct *affective quality*. As such, the process of aligning low-level resources with Poynton's dimensions followed multiple overlapping paths. For example, *power* can be found in the lexicogrammar in the systems of both MOOD and modality (as declarative statements for example) while manifesting in the visual modes, through the combinations of direct address, physical distance, and vertical gaze angles. Whereas deriving field's variability was a deductive process and suited to a grid-like approach, analysis of tenor required a more inductive approach, working from the low-level coding in MAXQDA.

Following O'Donnell (2020), Tenor was considered as a dynamic phenomenon, varying between the kinds of represented speakers and situations. To cater for this, I broke the analysis up into *global* attributes of tenor, and *varying uses* of tenor. Music, special effects, and general patterns in the verbal mode were considered as global, rather than reiterating them for every instance found in the text. The texts were then considered for each participant's contribution to tenor.

MODE: Finally, an analysis of the MODE variable was used to draw out the characteristics of CVR in two overarching ways: first, as a semiotic technology and mode of communication and second, for the way language is used in CVR within its technical and semiotic constraints. MODE was analysed through the four variables listed below (cf. Eggins, 2004; Matthiessen, 2009).

- Channel: the sensory modalities involved in communication
- Bandwidth: the directionality and richness of the flow of information
- Medium: the written-ness or spoken-ness of a text
- Experiential distance: how language intersects with the phenomenal world

For first material aspect of MODE, I analysed the MODE variables of *channel* and *bandwidth*. Analysis of channel targets CVR's immersive materiality as a series of overlapping sensory channels (e.g., visual, aural, tactile, proprioceptive). Here, CVR is examined non-linguistically for the way that Information flows across the expression plane, and how the spectator is brought into contact with CVR texts on a sensory level. The prototypical face-to-face, conversational mode of communication was used as a 'baseline' for appraising bandwidth, against which the texts were compared.

The second linguistic aspect of MODE was analysed along the dimension of *medium*, used to probe the texts for how they operate linguistically within CVR's channels of communication

(technologically asynchronous, one-way, and monologic). Medium is realised in text by the lexicogrammar as a dichotomy of 'written-ness' and 'spoken-ness'. Medium thus allows for the texts to be described lexicogrammatically for the ways they 'use' immersion and presence linguistically to engender naturalistic, dialogic encounters. Table 4, below, adapted from Eggins (2004, p.93) shows the variability of medium. The properties included were ascertained through linguistic properties such as lexical density, grammatical metaphor, general complexity (written), as well as grammatical simplicity, continuatives, personal references, and deixis. This phase of the analysis was aided by an online lexical analysis tool *Analyse My Writing* (Anon., 2018), with the results tabulated in Excel into categories, following Ure's categories of lexical density as relating to register (1971; see also Castello, 2008): from sparse to dense - spoken, fiction writing, exposition and technical density. The full list of shots and their lexical densities are in Appendix G: Lexical Density Analysis.

Face-to-face Spoken Interaction	Written
Dialogic (turn taking)	monologic organisation
Context dependent	context independent
Structure: dynamic (on the fly)	synoptic (comprehensive)
Spontaneity	final draft

<div align="center">Table 4: MODE's medium variable, adapted from Eggins (2004)</div>

Following this, I analysed the texts for the way that the immersive scenes intersect with the phenomenal world, as the 'experiential distance' between language and action. Experiential distance is used as a frame to describe the flow of information (as language and action) in the many configurations of speakers as *seen and heard, heard but unseen, seen and heard in diegesis*. Broadly speaking this dimension of mode allows for language-use to be scrutinized along the axis of action-reflection. language can be *embedded in action* thus playing a 'support role'; language can be *constituting action* in which case there is no 'action' to be found except as instantiated in the utterance. As with medium, ED is described partly the for the way that texts *simulate* experiential distance to meet the producers' immersive aims, where speakers construct their speech to make it appear that they and the spectator are together, jointly processing the scene around them.

AXIS OF ACTION > REFLECTION		
Action ⟵───────────────⟶		**Reflection**
Language embedded in action: where language is enmeshed in an action	Language commenting on an action: where language is supplementary to action	Language reflecting on action: where language alone constitutes the experience of external actions

Table 5: Axis of action and reflection, from Eggins (2004, p.91)

4.4.4   CVR's Ranking Order

Finally, studying CVR's modality required a preliminary set of 'grammatical units' to be used in delineating other functional units (individuals, groups, locations for example). This corresponds to the system of rank in SFL where phonemes constitute words, which form clause-complexes, and so on upwards through clause-complexes and sentences. Establishing CVR's rank provides the analysis with a set of units to manage the complexity of CVR. The CVR image exists essentially in two phenomenal forms: the field-of-view (field-of-view) that the spectator is currently choosing, and the 'sphere', i.e., the sum of all field-of-views that were captured by the camera at that moment. As a unique modality, CVR has a unique visual rank, where its smaller units are nested within larger units. In the previous chapter (section 3.8.1 above Rank) I described the two systems of rank that are relevant to this study, O'Toole's (1994) visual artworks rank and O'Halloran's film rank (2004). They are reiterated below for context. I will make a case here for a conceptualizing of CVR's rank by drawing from these systems and contextualising them within (non-linguistic) contemporary VR concepts that address similar concerns. The ranking order in the painted image, as conceived by O'Toole is as follows:

- The work (such as the painting)
- An episode (a grouping of people/things),
- A figure (distinct person or thing)
- A member (a part of a figure)

O'Toole's system was extended by O'Halloran for film (2004, p.117) along similar lines to O'Toole but reflecting the temporal nature of film and the interaction of the camera in the scene.

- The film plot
- The film's sequences
- A film scene (multiple shots)
- mise-en-scéne (a single shot potential)
- The frame (the specific angle that instantiates the shot)

89

O'Halloran includes O'Toole's lower ranks *figure* and *member* included also. Regarding their potential input into CVR's ranking system, the relationship of O'Halloran's *mise-en-scéne* to *frame* is formative, as she includes the former as a notional entity that is not visible in its entirety and rather is the overall cinematic space that is revealed through the frame.

CVR can be described similarly with a 360° mise-en-scene (what the camera captured), and with the spectator's current field-of-view equating to O'Halloran's frame. The difference being that the spectator selects the frame, where the director and camera operator would in conventional film. Rothe and Hussman (2019) make a similar conceptual differentiation between the *space* and a *shot* where the former equates to the total image and the latter is the field-of-view. Unlike framed film, CVR's equivalent to O'Halloran's mise-en-scéne is less of a notional entity, as it is always perceptually available to spectator, and in this regard, it is also like O'Toole's painted work.

Where O'Halloran's *frame* can isolate the lower ranks of *figure* (medium shot) and *member* ('cut-in') and present their contents sequentially, CVR is more like the painted image as it allows for episodes, figures, and members to be potentially observed simultaneously at all rank levels (although unlike painting, not necessarily in a gestalt). Here the primary difference lies between her *frame* and its approximation in CVR, which most closely resembles the spectator's field-of-view. As such, the only alterations that are required from the systems are to establish a rank equivalent to O'Halloran's Mise-en-scene and to adjust the terminology for easy comprehension in the discussion. As such I have opted to rephrase *mise-en-scene* as 'shot' (in other words, what the camera captured) and *frame* as *field-of-view*, while reserving the use of the term mise-en-scene for the objects that are contained within the scene. I have summarized these key differences in Table 6, below.

CVR's Visual Ranking order Compared to O'Toole and O'Halloran			
O'Toole's artwork	**O'Halloran's film**	**Proposed CVR**	**Rothe and Hussman**
	Plot		
	Sequence		
	Scenes		
Work	Mise-en-scene*	**Shot**	*Space*
	Frame	**Field of view**	*shot*
Episode		**Episode**	
Figure	Figure	**Figure**	
Member	member	**Member**	
*Mise-en-scene is reserved for us in describing the contents of the shot e.g., props, buildings, settings			

Table 6: Proposed ranking system for CVR, extending O'Toole (1994) and O'Halloran (2004)

4.4.5 Concluding Remarks

This chapter outlined the approach that is taken in this study, how I used the social semiotic frame described in the previous chapter. As a novel low-level study into CVR's discourse it was necessary to outline the challenges inherent in studying a novel form that resists many of the practices that exist for studying more mature forms. I addressed this from the data collection, software choice, transcription methods used outlining a potential path for CVR analysis.

On a deeper level, I put forward a means of integrating and extending extant semiotic systems in the study. Using SFL frameworks I constructed a coding system that allows the texts to show how their meanings manifest across SFL's strata. As a flexible system, the coding also allows for the texts to show where CVR's kinds of meanings push against, and extend the semiotics of the forms it remediates. This is key to understanding the unique ways that CVR configures discourse. The chapter culminated in a proposed ranking system for CVR, a necessary updating of extant ranking systems, through which CVR's complex semantics can be more accurately articulated.

5. Construing Experience

This chapter describes CVR as a semiotic technology in systemic terms. Specifically, the texts are probed through the ideational metafunction for the way representational meaning is configured in the 360° image, and across modes. Without a rectangular frame to manage the inclusion and exclusion of visual entities, CVR presents a heterogeneous field of activity that presents phenomenally as *simultaneity* and *multiplicity*. The spectator's rotating engagement is described for its implicit role in the construction of meaning as rotation constructs meaning relating to participants, processes, and circumstances. Findings discussed here therefore address the study's first research question regarding the spectator's role in the configuration of CVR discourse. The chapter also introduces the wider themes of the study relating to impact of CVR on the nonfiction spectator and producer. I will argue that CVR is semantically dynamic at the metafunctional level, where the spectator has a degree of semiotic autonomy in the construction of text. This will form part of the larger thesis in the final findings chapter, where this autonomy will be framed as a significant part of CVR's claim to transparency, and unique access to reality.

The second research question is addressed regarding the authors of nonfiction CVR texts and the potential effects of CVR on production. The spectator's semiotic autonomy mirrors a semantic indeterminacy, where meanings are at risk in the 360° image in an unprecedented way. Multimodal meanings are described for the way the 360° image is used in conjunction with language leading to issues regarding semiotic economy for the producers of the texts. Using language to support the immediate 360° space results in semantic redundancy requiring strategies to leverage the spectator's immediate experience while also articulating deeper exposition.

5.1   Experiential Metafunction: Visual Participants and Processes

CVR merges and complicates experiential readings of participants, their processes, and circumstances (PPC). To shed light on CVR's representational novelty, I will discuss several configurations of PPCs from a transactional and conceptual perspective. Transactional processes are presented at different ranks to account for the way transitive meanings are 'available' to the spectator across the shot and how the choice of field-of-view has experiential and logical repercussions for meanings. Conceptual processes are discussed as problematic owing to the omnidirectional nature of the image capture which presents a ubiquitous *locational setting* and field of activity. This challenges conventional conceptions of 'non-transactional' imagery and forces a rethink of how circumstance contributes to conceptual meanings. Issues of circumstance pervade all the other aspects of CVR's experiential function

and is addressed specifically for this at the end of the section, where I suggest how CVR's complex visual relations give rise to more logically oriented enhancing circumstances.

5.1.1   Transactional Processes

In the simultaneously occurring field of 360° activity, transactions in CVR range from low-level processes occurring close to the spectator, to processes unfolding at greater distances and in opposite directions. As such, rank, as proposed in the previous chapter is required to describe these transactional phenomena. To capture CVRs experiential novelty, I have broken the discussion down along the following phenomena, each of which will be addressed in detail.

1. As low-level actions constituting higher level processes (figure-member).
2. As discrete parallel episodes occurring across the shot (figure and/or episode)
3. As mid-level transactional processes forming an explicit higher-level process (figure and/or episode)
4. As transformative relations whereby one process transforms and augments the meanings of another (figure and/or episode).

As low-level actions constituting higher level processes. Having no frame, CVR does not afford close- ups and cut-in shots and the total body is always available (barring occlusions) where rotation will give the spectator access to any front-facing part of a filmed subject. Also, with a fixed focal length and wide angle-of-view, the CVR camera's field-of-view will necessarily include a large amount of the filmed subject. An implication for this is that paralinguistic features are always perceivable in the act of speech, as low-level actions contribute upwards to higher ranks.

For context, a preliminary analysis was performed where subjects at varying distances in the texts were viewed in a static field-of-view, to gauge how much detail was available to the spectator at varying distances (See Figure 11, below). While these could only be considered as approximations, they were useful in equating distances to conventional shot-types to explore the differences between framed and CVR subjects. At the closest perceivable proxemic boundary, where the spectator is bordering between intimate-personal distances and not rotating their perspective (roughly making eye contact), the field-of-view equates to a medium shot, also referred to as a 'knee shot' as the view takes in the subject as far down as the knees. Mamie was used as a reference, viewed in M#2.[12] At a roughly near-social distance the field-of-view resembles a 'full-shot' such as Alastair Leithead in C#4 (named for taking in all the subject's body).

---

[12] These are based on the Oculus Gear VR HMD with a horizontal field-of-view of 100°

At the longer proxemic boundaries of social-public ('long-shots'), the entire subject is visible and relatively small in the view, such as in M#13. Even at intimate distances most bodily actions, from the waist up are perceivable in the shot, meaning that even prior to a rotation, there is little potential to isolate faces, or other paralinguistic information.

Mamie at approx. intimate distance (0:41)

AL at approx. personal distance (1:02)

Mamie at approx. social/public (4:46)

Figure 11: proxemics visualised as shot-sizes

The implications of this can be viewed in several ways. From the perspective of semiotic autonomy, the spectator is essentially free to provide their own cut-ins, reflexively contextualising speakers and adding paralinguistic information at lower ranks. The implications for the producers of CVR images is the semiotic control of these low-level processes, and the higher-level meanings they constitute. For example, Mamie in her first shot is absent-mindedly dipping her hand into the water, bringing her hand into relative focus. Whether or not this negatively impacts the overall conceptual meaning of the shot (e.g., she is visibly *relaxed* in her movements), it signals that the lower rank (member in this case) is always 'at risk' of contributing to higher-level meanings.

Alternatively, the effect can be characterised as positive, such as in the later shot (M#7) where her grooming activity is replete with visual, verbal, and kinetic actions experienced simultaneously. Here the low-level actions are perceived with minimal rotation but nonetheless allow the spectator to decode the actions, gaining a degree of semiotic agency that is contrastive to being led, for example through a framed continuity edit. Tricart (2017) refers to this semiotic autonomy where storyteller 'offers a "space" which is set and staged and lets the audience do the rest' (p. 85).

Discrete parallel episodes occurring across the shot: In this configuration, processes were found to co-occur in the image without their necessarily contributing to explicit higher-level process.[13]

---

[13] It is always possible to resolve all entities at all ranks into a singular high-level process, such as

Unlike the first type, they concern mid-level actions (e.g., playing, cooking) that can be spread between figures and episodes around the shot. These occurred primarily in shots with lots of inhabitants, such as M#9, where Mamie is surrounded by episodic units such as the girls reading, and those adjacent to her on the wall, and the woman sitting nearby cooking. Where there is no explicit transactional involvement between processes, and particularly with Mamie, the principal participant's processes, they become auxiliary and circumstantial. Instead, processes build more abstract higher-level circumstantial meanings for Mamie, beyond that of location, interpretable loosely as *accompaniment* (*with* friends, neighbours, social equals). Where individual processes become grouped together as having a homogenous input into meaning, there are resonances with the linguistic effect of 'down-ranking', where participants in clauses become absorbed in noun-groups (Halliday and Matthiessen, 2013, p.10). Downranking in CVR is experienced dynamically through rotation, where entities are categorised and arranged into units, in real time.

The third and fourth configurations share similarities, each involving multiple processes contributing to a high-level activity. They differ in terms of the kinds of things represented, as either explicit activities (teaching and learning, church celebrations, train station greetings and farewells) or more implicit and conceptual kinds of meanings. *Mid-level transactional processes forming an explicit higher-level process* is explicit in the way it configures processes. Here activities were often supported by the setting, either institutional or social. In M#6 for example, the pupils are seated at their desks, sitting, gazing, waiting: an adult enters (2:01), and the pupils' gaze becomes part of a coordinated higher-level *teaching* process. Their process becomes absorbed into the larger process and because there is a causative aspect to this, it is also logical in nature. As such, Martinec's logical conjunctions can be applied here also to describe the dynamics involved. Here, the transaction-based conjunction involves the teacher *coordinating* the action vectors of the pupils (Martinec, 1998, p.163).

The fourth kind of configuration, *transformative relations* allows field-of-view to be factored into the connecting of participants and their processes. The effects are 'transformative' in that processes in each field-of-view are enhanced in some way by another. By connecting disparate processes, the role of the participants is altered, and the sum of both processes produces more complex meanings at the higher rank of shot, when connected by the rotating spectator. Visual Grammar accounts for this kind of transformation as a *conversion* (Kress and van Leeuwen, 2006), where for example an *actor* in one process can simultaneously be the *goal* of another co-

---

'socializing' or 'education', but this would be reductive and counterproductive at this level to understanding the more complex aspects of CVR's transitivity. These higher-level activities will be a feature of *field*, discussed in the register analysis.

occurring process. A distinction here is that when the processes in the field-of-views become larger processes they tend to have more complex logical and conceptual meanings for their comprising multiple 'nested' processes below.

Mamie's Dream provides two shots that illustrate this phenomenon. M#3 presents a conversion, and M#18 presents a conversion with complex logical meanings. The conversion in M#3 can be described in grammatical terms. The shot begins with two alternating field-of-views each comprising distinct episodes engaging in mid-level processes (footballing and spectating) that combine to form an explicit shot-level activity of *watching a game.* The spectator's initial orientation will determine the roles of the participants involved where depending on the field-of- view, the footballers or the watchers will be seen first, forcing a temporal order of sorts onto the ideational construction. In other words, *the boys are playing football - while being watched*, or *the pupils are watching - a game of football* (they are not 'football spectators' until the footballers are seen). As such, the footballers 'convert' the watchers not just in basic transitivity but also as a consequence of rotation. The underlying point is that there is a perceptually deterministic aspect to the order in which the spectator encounters transitivity, where the field-of-view cannot accommodate all involved. This complements the conversion process as laid out in visual grammar and adds the rotational field-of-view.[14]

M#18 illustrates this effect more sharply. The shot shows her and a large cohort of pupils engaging in a school celebration comprising verbal exchange, reciprocated gaze, and a strong synchronized kinetic engagement. In this regard the shot is like other shots where Mamie is engaged in processes with her pupils which do not transform each other significantly. Mamie's processes here differ as they are contextualized by the seated adults who share the stage with Mamie. The participation of this episode is arguably parallel (the adults do not jump, dance, or sing and are adjacent to the action). The processes connect and overlap however where one of the seated participant's gazes is extended in Mamie and the dancers' direction. Mamie and her own material processes become *mental phenomenon*; her role is transformed via the seated episode and their own presumed signification (as also on-stage and as figures of authority). Her perceived status is at risk here of ambiguous signification relative in one direction to the children and in the other to the seated adults who have presumed status.

The simultaneous space presents a degree of multiplicity and indeterminacy regarding transactional meanings. The configurations presented here, and particularly the *transformational* kind suggest that the rotating spectator has the capacity to decode logical,

---

[14] This could also be considered as reflected in the other functions as alternating *Mood-residue* and *theme-rheme.*

enhancing meanings as they scan the shot. Transactional processes such as in Mamie's case give rise to conceptual meanings, which are also circumstantiating meanings. Thus is the complexity of analysing CVR's transitivity, which I will elaborate in the following subsections.

5.1.2   Conceptual Processes

CVR presents similar challenges in finding clear examples of conceptual image structures such as those used in analyses of framed imagery. Examples in conventional image analysis tend to present conceptual participants as decontextualised from their narrative settings and presented against neutral or abstracted backdrops, as in Figure 12, below (Kress and van Leeuwen, 2006).[15]

Figure 12: Conceptual image from Reading Images (Kress and van Leeuwen, 2006)

Participants in such configurations are scrutinised for their internal qualities, and where multiple participants are presented, they are logically related to each other. This is relatively difficult to achieve in the kinds of naturalistic representations found in CVR, where the setting is always visible and often with participants engaged in transactional processes. Shots are at best considered conceptual by the main subject lacking distinct process vectors. In the case of human participants, this is problematised by the directional nature of gaze. A framed image can present a person passively looking 'out of frame'; in 360° there is always a visible target of their gaze unless they are looking into an occluded area (out of windows for example – discussed

---

[15] For other examples see van Leeuwen and Jewitt (2001, p.139) and Liu and O'Halloran (2009, p.374).

as *Framing* in chapter 7, section 7.1,). Analysing conceptual processes in the texts involved looking for images that skew away from overt transactional processes. The following examples show how conceptual image techniques were employed in each text, as *possessive*, *identifying*, and *symbolic- attributive* processes. The texts present different kinds of conceptual images, exemplifying the contrasting styles adopted by the texts and the nature of the subject matter. *Congo VR* is relatively straightforward in its use of non-transactional imagery when it is presenting images with no human participants. *Mamie's Dream* has no such shots and is more inventive in establishing conceptual meaning.

Congo VR uses conceptual processes in two principal ways: in the form of aerial shots (#1, 2, 3, 5, 12, 16 and 23) used to present the visual attributes of the DRC, and on ground level where Mobutu's ruined palace provides a conceptual accompaniment to the historical voice-over (#24, 26, 27, 28 and 29). The opening shot comprises a generic series of landmasses and bodies of water. In this case the shot in toto is carrier, and its relational attribute approximates to 'vastness'. The visual scene is also a *possessive attribute*, glossed as 'this landmass has large expansive body of water'. Continuing the aerial descent in the subsequent shots, the vistas continue to present an array of possessive-attributes ('*has* far-reaching and dense forestry'; '*has* industrial markings'). The later shots in the decayed remains of Mobutu's palace are devoid of human actors. Attribution here is related largely to the building's decay in shots by contextualizing visible remnants of the palace (pool, fountain, cross filled with water) within the decaying structures. Identifying relations of token-value are possible in many of these images. The remains of the swimming pool (C#27) for example, are token, the 'swimming pool *is ruined*' relationship is extended by attribution of identity, where swimming pools *are luxury things* and as such 'this ruined swimming pool *was* (*a rich person's pool, a site of recreation*)'.

Mamie's Dream is more ambiguous regarding conceptual processes. When analytical meanings are considered possible in the absence of discernible transactional actions, *Mamie's Dream* was found to clearly realize such analytical meanings in three of the nineteen shots (#2, 8, 19). In the three shots in question, Mamie is static and not engaged in any activity either as actor or goal (it is difficult to exclude mental processes and target of gaze, as looker or looked-at). As non-transactional, she is therefore 'at risk' of conceptual signification but in all cases her visual attributes, minus the verbal accompaniment can at best be assigned as 'alone' or 'thoughtful'. Her visual circumstances must be factored in for conceptual meanings to become more nuanced. In M#2, Mamie is at a distance to the river below and to the landmasses at either end of the bridge. Her gaze is semi-focused, and its target is the horizon. As shown in Figure 13, below, the *bridge-as-circumstance* and the resulting distance to the generic human participants (with whom she does not interact) results in her not just being 'alone' but also expanded to mean 'away from'

visible social entities (in blue box), with implications for the extent of her alone-ness ('extent' also being a circumstance type, discussed below).

Figure 13: Mamie's isolation expanded by other social actors (0:46)

Shot eight (Market) differs with respect to the use of circumstantial features. Once again, she is isolated from her surroundings spatially but with additional effect of *temporal* isolation where frame speed has been manipulated to exaggerate all but Mamie moving at faster speed.[16] The effect lends an ethereal aspect to Mamie's isolation and the relational-attributive process is Mamie as carrier, 'out of sync' with the market dwellers, isolated, lonely, depressed, all as potential attributes. In visual grammar, this would also be considered a *symbolic attributive* process, where the environs impute an attribute (slowness) onto Mamie, also describable as internal processes *symbolic-suggestive* through her physical demeanour where 'meaning and identity [are] coming from within' (Kress and van Leeuwen, 2006, p.68). In other words, using visual metaphor terminology, meaning here is achieved partly by transferring the *source* in the form of the overlaid fast movement, onto the *target* (static Mamie) to achieve the symbolic meaning (see Feng and O'Halloran, 2013). When the circumstances infer such relational meaning on participants, the logical function must also be considered for its contribution. Here *simultaneous contrast* enables the conceptual processing: 'she is slow *compared to* the fast market people' (logical) becomes 'Mamie is unlike most people' or 'Mamie is an outcast' (relational identifier).

The images in the texts can also be considered for their fusing of transactional and conceptual processes. Kress and van Leeuwen discuss how transactional processes in 'naturalistic' (in other words, 'figurative') images will necessarily contain "embedded 'analytical' processes" relating

---

[16] Mamie is likely also sped up, but it is not perceptible as she remains almost completely static throughout.

attributive details pertaining to participants and circumstances (2006, p.50). While this stands true in CVR, the reverse can also be found in a manner perhaps unique to CVR where *transactional* processes at the rank of episode are contained within *analytic* processes at the rank of shot. M#9 is a case in point, where its abundance of low-level transactional processes; reading, playing, cooking, are 'building' upwards to a shot that has no high-level vectors (they are all circumstantial as *accompaniment*). As such, the shot has a conceptual dimension to it where Mamie is carrier to an attribute of 'sociable' or as part-whole relationship as 'part of the group'. Unlike Kress and van Leeuwen's example, the transactional here is subordinate to the conceptual, a feature of the omnidirectional image and its resulting simultaneity of its low-level processes.

The encoding of processes in the CVR image is semiotically fluid when compared to framed images. Transactionally, participants' statuses and roles are dynamic and depend partly on the order in which they are read. The four transactional phenomena described here can partly account for CVR's novelty as a representational medium, describing the interaction of transitive elements across rank. A common thread across all appears to be the 'logicalizing' of meanings as processes are related to each other upwards towards the rank of shot, just as clauses realise meanings in the larger context of clause-complexes. The spectator's natural tendency is to relate things to each other to make sense of things. This will be addressed directly in sections 5.1.4 and 5.2 of this chapter, dealing explicitly with circumstance and logical relations. It is mentioned here for the way the logical permeates the transactional, in the CVR context. The CVR camera has a direct impact on representations of conceptual processes. As discussed, it is difficult to render the kinds of images used in visual grammar analyses without recourse to framing and other means of decontextualising the participants. It is also perhaps a fundamental aspect of CVR that its representations will always skew towards narrative where presence, in and of itself, implies a kind of 'meta-setting', where narrative is always imminent and where the spectator's reflexive movements add a 'meta-transactional' component.

In the discussion thus far of conceptual images, I have intentionally skirted around two factors. One is the omnipresent drone in *Congo VR*'s aerial shots, which disturbs the conceptual nature of the shot. It is highly kinetic and implies both a kinetic vector and a notional actor-goal relationship, where the spectator is 'carried' or 'flown' across the scene. To read this as transactional would fundamentally alter the shot and presumably negate the preferred reading. I consider this to be an erroneous use of visual resources from a transitive perspective and will discuss the drone in more detail as an interpersonal dimension to the production, lending the spectator a problematic of point-of-view (section 6.2.4, Technical Perspectives). The second issue is related to some of Mamie's conceptual shots, which would be intuitively considered as *mental* processes, where she is thinking of something that is most likely articulated in the voice-

over (we hear her thoughts). While this is a natural cognitive link to make, combining both into the most likely event, it is difficult to support from a semantic perspective. I will return to this briefly in discussions of the logical functions of such shots, regarding the logic-semantic conjunction of projection.

5.1.3   Experiential: Circumstance

Images and linguistic clauses differ fundamentally regarding circumstance. Speakers can choose to omit circumstance, where the producers of camera-based images cannot. The CVR camera's optical system compounds this fact as aspects of the visible setting are not just present but visible in sharp detail, and where circumstance is permanently at-risk. Compared with SFL's library of circumstance types and subtypes (see Halliday and Matthiessen, 2013, pp.310–320), visual grammar posits just three kinds: *setting* (location), *means* ('by means of'), and *accompaniment* (Kress and van Leeuwen, 1996) which were used in a visual analysis of the texts. Circumstances of *means* fits CVR's semantics without the need for reappraisal. *Setting* is problematic in CVR's inability to exclude it. *Accompaniment,* as defined by Kress and van Leeuwen also becomes problematic. In visual grammar, a simplistic distinction is used to distinguish between a participant and an accompaniment where entities are considered accompaniment by virtue of their not intersecting with transactional processes - in which case they would have participant status. I will discuss visual grammar's circumstance types in the context of the texts and discuss the ways that CVR both complicates established conceptions and suggest how CVR also accommodates different kinds of circumstantial features.

As I have discussed, *place* is always available in CVR in a detailed manner. As such, the films necessarily use the setting to produce contexts for the films' participants processes. Here the films diverge in their use of circumstance. *Congo VR* explicitly forefronts the setting as its principal subject matter, at times just presenting place, such as in the aerial shots, and those in Mobutu's palace. Here, place has a semantically ambiguous quality, where participants (in the general human/non-human sense) and circumstance merge. The sunken waterfilled cross in C#28, and the taxi in C#11 are both participant and circumstance. Elsewhere, location is used as the 'local' context for each of the social actors and groups, showing them in-situ (e.g., Junior and Mama at home ('I live here' C#13); Sapeurs on the dancefloor C#15 and Pygmies in the forest). Mamie uses setting in a similar manner in places but as I have described previously, she also uses setting to provide her with symbolic attributes, used to enhance the co-occurring verbal narrative and its processes. Her locations are variously in *nature*; in *social/public spaces*; and in *school*. In this regard, setting attributes her with qualities relevant to the different phases of her personal history: before she is evicted from the family home; her period of isolation

before the learning assistance programme; the contemporaneous post-LAP phase of her development.

Where other participants are visible but in parallel processes, circumstance of *accompaniment* (as per visual grammar) is feasible. Kress and van Leeuwen qualify accompaniment as: "participants which could be left out without affecting the basic proposition realized by the narrative pattern, even though their deletion would of course entail a loss of information." (2006, p.72)*.* This is perhaps more straightforward in the context of still images, where the visible relationships are fixed in time, unlike CVR where movement in the 360° space make distinctions problematic and where participants become circumstantial entities and vice-versa. There is a clear distinction in *Congo VR*'s shots between the primary subject(s) and the more auxiliary figures who are circumstantial accompaniments. In *Congo VR*, the pygmy tribesmen split into arrangements of actor (man digging C#19; pygmy elder C#20) and sayer (pygmy elder C#21), with all others watching but not interacting in their processes. Mamie presents a similar dynamic in M#13 where her walkthrough of the busy social space is accompanied by many others, some watching, some engaged in tangential activities (playing 'keepy-uppy' with football), and some seemingly oblivious to her presence. C#18 differs in that it presents the tribesmen as swapping roles between actors/behavers in the enacting of material processes (searching/hunting) and circumstance of accompaniment, depending on the portion of the shot being viewed.

Novel phenomena emerge when the CVR image is considered for enhancing and elaborating circumstances. Enhancing circumstances take the form of *contingency*, as 'under what condition' (Halliday and Matthiessen, 2013)), and the elaborating circumstance, of *role*, a lexical example being '*as* your teacher, I will…'.[17] Novel circumstantiation emerges where sequential vectors of action connect participants for portions of a shot and when they cease, their lingering effect serves to contextualize and circumstantiate others. Essentially, processes when perceived over time 'generate' circumstances in a dynamic fashion, and vice versa. A shot that demonstrates this is M#16 involving a group of pupils dancing in the school yard (5:41), seemingly at the behest of an adult male who appears to be in a supervisory capacity. There are simultaneous and sequential processes occurring, involving figures that alternate between participants and accompaniments. The shot is deconstructed below as a series of 'beats', or sequential low-level transactional processes contributing to the overall action. The transactional processes involve the adult male instructing a group of girls (ep.1, below) to commence dancing. Having done so, he disconnects his action vectors and walks away. In this regard, he has two circumstantial

---

[17] Apart from the example of a social role given here, non-human kinds of roles are also part of Halliday's schema e.g., 'set the record *straight*' (Halliday and Matthiessen, 2013, p.243). As these appear to be purely lexical, they are excluded from analysis.

features, *means* and *accompaniment*. His physical gesture is circumstance of *means,* in other words 'communicating *with his hands'*. And whereas he begins the shot as a participant connected by vectors to the other episodes (ep.2 and ep.3) he ends the shot as circumstance of *accompaniment*, no longer connected to any other participants. To borrow from SFL, he is *rank shifted* from process to circumstance. When contingency and role are included in the analysis, the following additions can be made to M#16. First, having been prompted to dance, the girls in ep.1 are enacting behavioural processes, with the agent-medium discontinued. They now have a circumstantial feature of 'under instruction' where their dance takes on a quality in its being contingent on the man's previous verbal process. Regarding role, the man's actions infer on him an authority to give instructions, which carries connotations of his *role* in affairs. In this shot alone, he is 'supervisor/instigator' of sorts, as any more detail than that would require either cross-modal signifying or conjunctions with other shots (for example, 'before, there are [were] only *male teachers*' in the preceding shot M#15). The novel aspect of this is that the circumstantial features are transformed in-shot and perceivable simultaneously where the act of instruction circumstantiates both he and the dancing girls.

M#16 Girls Playing. Broken down by beat, processes, participants, and circumstances

1. Three groups of pupils, standing, dancing, and playing (5:35)
    a. Pupils ep.1 standing/watching man in white [senser/behaver]
    b. Pupils ep.2 dancing adjacent to the upcoming activity [behavers]
    c. Pupils ep.1 playing, peripheral to the main activity [actors/goals]
    d. Man, verbally interacting with ep.2 [sayer]
2. Man begins verbal process through 'off you go' gesture (5:40)
    a. Pupils ep.1 are target of his gestures [goal/medium]
    b. Pupils ep.2 and 3 are unaltered **{are circ: accompaniment for ep.1}**
    c. Man causes dance to begin [agent] **{circ: means (with hand)}**
3. Girls begin dancing – Man walks away (5:41)
    a. Man ends material process, **{circ: accompaniment/role}**
    b. Pupils ep.1 are dancing [behaver] **{circ: contingency}**
    c. Pupils ep.2 become circumstantial for Pupils ep.1 dancing *with* ep.2 **{circ: accompaniment}**[18]
    d. *Pupils ep.3 have ambiguous circumstance. Can be contiguously temporal in other words, 'while the girls play'*

In addition to the main process, there is a larger circumstantial arrangement taking place in this

---

[18] It is also possible here that ep.1 and ep.2 have become a single episode through their newly combined process.

shot as illustrated above where the playing pupils ep.3 never interact with either of the other two episodes and are thus 'twice-removed' from the primary actions. In this case, it is feasible for them to contribute to a contiguous temporal circumstance on the dancing girls akin to 'while other's play nearby' but should be done tentatively to avoid overburdening the analysis and lexicalizing the imagery unnecessarily.

5.1.4   *Circumstance: Concluding Remarks*

The lack of framing in CVR makes circumstance a ubiquitous yet unstable entity. The circumstantiation described here points to the fuzzy boundaries in the grammar of CVR, where processes and participants are imminent circumstances also. The 360° photographic modality incorporates a detailed setting, writ large in all directions regardless of how much the producer wishes to decontextualise their participants. The setting's ubiquity means that it is always at risk of infiltrating the other semiotic choices, particularly where conceptual meanings are to be attributed to participants. Without a frame therefore it is a requirement to move the shot to a place where circumstances are less likely to interfere with the main topical processes, which would result in an editorialising of the 'reality' captured by the camera.

Circumstance of accompaniment should be read as dynamic and in flux, where participants come together in joint processes, disengage but crucially – remain in the shot. Whether an entity is participant or circumstantial is dependent on which process is currently considered to be 'the basic proposition', and as the example above shows, this can be in flux. In this regard, the field of activity becomes a logical space, where processes leave circumstantial enhancing 'imprints' of sorts on the unfolding actions.

5.2   Logical Metafunction: Visual Conjunctions

The previous sections described the logical aspects of the 360° image, as an experiential, circumstantiating component. This section addresses logical meanings explicitly by applying extant visual theories to the texts. Using Martinec's action (1998), and van Leeuwen's filmic (1991) analyses as a guide, the texts were analysed for meaningful visual conjunctions, across 'conjunctively relatable units' (CRUs) (ibid.). The findings discussed here outline the low-level establishment of visual logical meanings. The subsection deals with fewer semantic properties than the experimental function and is as such brief. The value of the logical metafunction the visual analysis of CVR is in the way it is used with experiential meanings, to expand and conjoin them in larger complexes of meaning. As such, this section is formative for the larger macro analyses of logical conjunctive meanings, in chapter 7, section 7.4.1.

Where logical conjunction occurs between CRUs of equal rank, they can be compared directly for

simultaneous contrast (Martinec, 1998, pg. 166). The effect of similarity coalesces the figures into their episodes. For example, the dancing sapeurs in C#15 comprise an episode of three figures who move in synchronicity and are connected by their kinetic similarity. The BBC team are a co-occurring episode who are also internally similar in their attributes. At the rank of shot, however the two episodes form a comparative relation of difference (or *negative comparison* as per van Leeuwen) where the BBC team episode are different to the sapeurs in their attributes and processes. The spatial component, where each episode inhabits opposing sides of the space, means that the spectator resolves these differences through rotation. A similar effect is found in *Mamie's Dream* where the footballers in M#3 are internally *similar* as kinetic and fluid but *different* compared to the static and ordered watchers above.

Distinctions become less clear where contrasts are found across varying ranks such as in M#5 where Mamie can be considered contrastive to the playing girls, but where she is *figure* to their *episode* (as she is with the seated adults in M#18). At the higher level again, Mamie is found to be contrastive to the overall shot in M#8 where she is *alone and slow*, to the market scene's *social and fast*. Likewise, Alastair Leithead is found to stand out in C#20 being the only white skinned person in the shot. When the examples are compared, the contrasts are augmented by the shift in rank difference. Where the sapeurs and BBC (same rank) are more directly comparable, and Mamie is circumstantiated by her surroundings (higher rank). The difference is subtle but meaningful as a 'direct comparison' sets up a logical relation that is qualitatively different to a broader circumstantiation. As I have already described, however, the distinctions can be hard to uphold in the 360° image.

The conjunctive relation of *coordination* involves one action enabling another action of the 'same status', where they 'contribute equally to determining the development of the larger action sequence in which they participate' (Martinec, 1998, p.165). Martinec classifies 'the same status' being processes that are both transactional and both non-transactional. This is clear at the very low-level process-complexes that have knock-on and/or reciprocal vectors. Transactionally, the footballers in M#3 exemplify a coordinating action in their kicking, receiving, and returning the ball contributing upwards to the higher-level contrasts of *footballing* and *watching*. In a non-transactional fashion, the man dancing with his back to Mamie is *coordinated* by the band's process, as well as the wider reciprocal contribution of the congregants to the higher-level celebrating. The example discussed earlier from M#16 (the girls instructed to dance) is perhaps ambiguous as to whether it constitutes coordinating actions: the higher-level activity is essentially 'coordinated dancing', yet the man's gestures are a verbal process, whereas the girls dance process is quasi-material/behaviour.

Simultaneous contrasting and coordinating conjunctions differ significantly in their contribution to meaning. Coordinating conjunctions, such as those seen at low-levels such as the footballing actions (kicking-running) are less analytically useful than the mid-level conjunctions they contribute to. Similarly, the dancing sapeurs are interesting largely as an episode ('the sapeurs') rather than in their individual motion vectors, turns and gestures, which can as be glossed as 'synchronised dance', without affecting the meanings of the shot. Martinec acknowledges this, stressing that the level of detail chosen to analyse is subjective and based on the research objective. *Simultaneous contrast*, as evidenced between the sapeurs and BBC team in C#15, is of more interest to this analysis. It provides the grounds for logical inferences to be made in tandem with the experiential components, where differences and similarities in the visuals become logical operators, switching, and redirecting logical meanings across modes and shots which become logical complexes. This will be discussed through a macro-analysis of logical meanings in chapter 7, section 7.4.1.

Taken as a whole, the web of visual conjunctions in the 360° image, either coordinating or contrastive, contribute to a characterisation of the form. All processes are available to the spectator as an interlocking system of low to high-level conjunctions. This points back to the simultaneity of the 360° image and the logical connection of information 'upwards' from low-level to high, as well as 'across' the interconnecting entities. Implicit in this is the immediate sensory awareness of difference and similarity as the spectator rotates their view.

5.3   Cross-modal Analysis

CVR discourse is considered here as a contextual phenomenon. High-level theoretical frames are used to establish the ways in which meanings emerge through the combinations of image and language. These frames effectively subsume the cohesive and conjunctive systems that will be the centre of analyses in later chapters. Here they are 'localized' as cross-modal instances of meaning within single shots. Of interest here is the way that these high-level theories describe the 360° image and voice, taken as the significant ideational portion of the modal ensemble.

5.3.1   Ideational Cross-modal Organisation

Taking a top-down approach, the image and the lexical content of the verbal track were considered for *parallels*, essentially as visual, and verbal similarities. Where parallels were found, the shots are discussed for *convergent* meanings; where no parallels are found meanings are considered *divergent*; and where only fragments of the image or voice is reiterated across modes, they are considered for *componential, instantiating* relations. Upon identifying instances of these constructs, they were then considered for their ramifications and the way meanings

could be described between modes.

The texts diverge regarding parallelism: a significant portion of *Congo VR*'s subject matter is represented both visually and aurally, with parallels emerging when the verbal and visual restate each other, whereas *Mamie's Dream* does not (and cannot at times) show much of what is in the voice-track, comprised largely of historic subject matter. These general trends notwithstanding, they each vary within their use of parallel structures, as found in the examples, below. The examples in Table 7, below on pg. 107 from *Congo VR* show polar uses of parallel structures, where the first shot has a full parallel (italicized in bold) and the other almost none. In the first example (C#19), the parallelism is clear as the transitivity is repeated with participants (*pygmy men* + deictic 'they('re)') and material processes (*striking thicket* + 'looking...') Here, each mode has equal status, and the dynamics are of *co-contextualisation* with a convergence of meanings that has the sole function of exposition (Gill in Unsworth, 2006). Elsewhere Alastair Leithead's speech diverges from the visuals in shots such as the second example (C#22), where the transitivity is starkly different between modes. The only common transitive component here is 'the people' which is an exophoric deictic reference, not referencing anyone directly in the shot but is a relation of part-whole or subtype relation (as hyponym).

Congo: 18 FOREST HUNT	
Immersive visual	**Voice-over**
Thickets of bushes and trees in all directions; one large fallen tree-trunk lays overhead. Men crouch and move slowly through, striking the thicket with small tree-branches. Man climbs into hollowed-out tree trunk'	*They're looking for animals and useful plants along the way*
Congo: 22 VILLAGE UNDER CANOPY	
Immersive visual	**Voice-over**
Interior semi built/covered filled with children and some adults. All attention focused on two men lighting a fire on ground. Fire lights and two men sit up to avoid smoke; continue to build fire with small objects.	It wasn't just the Belgians who exploited and brutalized **the people**. When independence brought a flicker of hope, their own leaders took the riches for themselves. and left **the people** with nothing.

Table 7: Parallel processes in Congo VR, shots 18 and 22

Mamie's Dream presents a similar dynamic in M#2, shown below in Table 8. In the voice-over, there is just one potential instantiation between Mamie, seen on the bridge with 'girls' as spoken by her father. Again, the connection is tenuous at best and other than that there are no parallels

to be found.

Mamie's Dream shot #2 ON BRIDGE	
Visual	**Verbal**
Woman stands on a quiet traffic bridge over river, hands clasped looking out along the river into the distance	My father said, "***girls***, they just get married and move away"

Table 8: Instantiation across modes in Mamie's Dream, shot 2

These shots form *instantiation* relationships (Gill in Unsworth, 2006) where one mode captures an instance of the other's broader transitivity. Both are tenuous examples of instantiation however, owing to the kinds of reference in the instantiating component ('the people', 'girls'). As such, there is a *re-contextualisation* between voice and visuals where the meanings of the visuals are changed by their conjunction with the voice. Much of the film's shots fall between these two extremes where varying numbers of transitive components are shared across modes. In Mamie #8 for example (Table 9, below), the visuals realise a higher proportion of the verbal transitivity where 'out onto the streets' creates a specific circumstantial parallel with the visualised streets, and 'I was alone' provides a conceptual/attributive match (as retiming effect and visual framing). As such, the shot converges in places (on streets, alone) and diverges elsewhere (father, prostitute, expulsion from home) realising *componential relations*. This is common across both texts, partly as a matter of style and partly of necessity and semiotic economy. Essentially, the long shot-durations present a singular extended visual and where the verbal track must 'fill in' additional information. This results in the image being a smaller component of the larger meaning resulting in an image with lower status than the verbal mode and is in a supportive mode (Martinec and Salway, 2005).

Mamie: 08 IN MARKET	
Visual	**Verbal**
Woman sits against a building facade out onto busy market. People sit, stand, and move about in the immediate vicinity (retimed)	My father said I was a prostitute, so he drove me and my belongings **out onto the streets**. **I was alone**

Table 9: Example of instantiation in Mamie's Dream, shot 8

Co- and re-contextualisation, as found in these examples can be described in the context of Barthes *anchorage* and *relay*, where in the hunt in C#18 the voice serves to anchor by *presenting* the meaning of the visuals (what the men are doing), and in C#22 the voice and visuals enter into

more of a *relay* where each extends the other (van Leeuwen, 1991, p.89) and where the image and voice create new meanings beyond those contained in either mode.

Martinec and Salway equate Barthes relay to the logico-semantic system of *enhancement* (2005, p.342) which is useful here in opening up descriptive potential of SFMDA for the relations that pertain in C#22 (the men lighting a fire). Here, enhancement provides logico-semantic meanings to be conferred on the visuals, where the single componential link of 'the people' affords the verbal track to become *causal*, in other words, '*independence brought a flicker of hope, their own leaders took the riches for themselves* [and because of that, these] *people* [were left with] *nothing'.* Of course, cross-modal enhancement needn't be as literal as this (we don't 'speak' these meanings to ourselves), but the re-contextualisation, even if vague, is nonetheless stark where those pictured, likely enjoying the fire-lighting spectacle, take on a different and more melancholic aspect. In Barthes rhetoric, M#8, when seen in the market differs from the tribesmen hunting in C#18 as her visuals are *illustrating* the voice by giving a visual snapshot of a larger set of processes. A similar cross-modal enhancement pertains here as she is seen partly as the *result* of what is heard.

A similar but more complex cross-modal arrangement can be found in C#17, illustrated in Table 10, below. Here, co-contextualization is formed through a strong cohesive tie involving three deictics ('this', 'they' and 'we') and where the voice reiterates and anchors the visual circumstance (Buyanger village) and process ('performing'). The logical connection here is largely a temporal enhancement where the voice circumstantiates the visual performance as 'before a forest hunt' but more consequentially an expansion of addition ties in the BBC crew part of the unfolding activities ('and we're going with them').

Congo: 17 VILLAGE TRIBE.	
VISUAL	**VOICE**
Village surrounded by tall trees, comprising thatched huts. Row of men moving rhythmically in tandem	**This is Buyanger village.** **They're performing a good luck ritual before a forest hunt.** And we're going with them.

Table 10: image-text breakdown in Congo VR, shot 17

All the preceding examples involved material processes in either the image, the voice, or both. Contextualisation structures and parallels pertain also to analytic meanings across modes. C#2 in Table 11, below for example, is a relational-attributive parallel structure. I described the image's conceptual transitivity in subsection 5.1.2 for the forest's attributes (*forest is far-*

reaching) and possessive-attributes (*has dense tree cover*). Here, the scale of the forest is reiterated ('stretch of rainforest') and enhanced in the voice-over ('the *largest* stretch of rainforest *outside the Amazon'*). The adjective group 'beautiful but troubled place' however recontextualises the visuals: 'beautiful' can be considered a form of visual collocation (by convention where, *forests are beautiful*), which is expanded by 'but troubled' as an *adversative addition* (van Leeuwen, 1991, p.80), adding a qualification to the forest below roughly as a 'shadowy unknown place where troubles lurk'.

Congo: 2 DRONE FORESTRY	
VISUAL	**VOICE**
Aerial vista of forest treetops, giving way to tree covered hills in distance. Cloudy sky with bright skies in distance	'It's a beautiful but troubled place, with the largest stretch of rainforest outside the Amazon'.

Table 11: Cross-modal conceptual meanings in Congo VR, shot 2

This points to Alastair Leithead's rhetorical strategy and the way he is using the image as the basis for his language choices. A strategy favoured by Alastair Leithead is to begin a shot with an instantiation between image and voice, used as a kind of 'launch-pad' for his more expansive meanings. This approach is taken in nine of the thirty shots,[19] where he anchors the spectator in the location from which point additional meanings proliferate outwards. A clear example of this is found in C#15 where he begins the shot with an instantiation of *sapeurs* (the visible dancing Sapeurs / 'The sapeurs display the energy'). At a componential level, he is moving from the *instance* outwards through a complex series of expansions threaded together with cohesive ties. Arguably, his verbalised 'displays the energy and brash confidence' is a parallel with the visuals but this is complicated by his voice attributing *qualities*. It is more useful to consider his use of metaphor as *comparing* the sapeurs to the capital (both having brashness and confidence). From here, he adds new albeit connected information; 'two-thirds of this country are under twenty- five'. The lexicogrammar, taken as a whole is threaded together through a series of collocations ('energy', 'confidence', 'young', 'burst free'); when the entire shot is taken as a cross-modal whole, the sapeurs are in one sense used to generate the far-reaching complex of social and historic exposition, while also retroactively instantiating it. He uses this strategy to good effect in many of the shots captured in Mobutu's ruined palace, where he anchors the spectator in the remains of his expanded historical descriptions, including the examples in Table 12, below (instantiated elements underlined):

---

[19] C#9, 10, 15, 19, 20, 23, 24, 27, 28.

Congo VR #28
That cross now **filled with water** is where his bed sank below the floor at night for safety and that paranoia was well-founded. He was removed by a coup in nineteen ninety-seven His empire of riches was left in ruins like that of his Belgian predecessors.
Congo VR #20
… rubber, it's **that white sap oozing out of the wild vines**. The pygmies were forced to search for them and to collect it

Table 12: Alastair Leithead's use of instantiation as a rhetorical device

Alastair Leithead's rhetorical flourishes are instances of logical argument, but they serve a more particular function in CVR, where location is imbued with the character of 'place illusion' (Slater, 2009). There is an aspect of distance involved, where Alastair Leithead is using the immediate space to bed his meanings. I will address this further at the level of register in chapter 7, section 8.1.3 (Experiential Distance) where his approach is a conscious playing with semiotic distance, as well as the informational manipulation of the spectator's field of activity.

5.3.1.1 Cross-modal Analysis: Concluding Remarks

Analysing the texts at this level is useful in determining the kinds of semiotic choices made by the producers within the constraints of the form and subject matter chosen. *Congo VR* exhibits a fluid alternating style between co- and re-contextualization, where the locative aspect of the film and its immediate processes are fore-fronted, instituting the spectator's presence as a semantic 'hook' for the language choices in the perceived 'here-and-now'. *Mamie's Dream* is constrained by its historic subject matter, requiring more instantiation and re-contextualizing meanings across modes. This variation in parallelism enables very different kinds of meanings. When *Congo VR* co- contextualises using deictic references and the reiterations of all transitive components, participants, processes, and circumstance, the modes contribute to a cohesive sense of contemporaneous experience (e.g., C#18 *'they're looking for plants'*). Recontextualising shots position meaning at a higher and more complex semantic level, as in Barthes *relay* where "text […] and image stand in a complementary relationship; the words, in the same way as the images, are fragments of a more general syntagm and the unity of the message is realized at a higher level" (Barthes, 1964, p.44).

The flipside of *Congo VR*'s co-contextualizing lies in the cross-modal redundancy (what is said, is seen, and vice versa) which raises issues of semiotic economy. Meanings that are co-

contextualising are redundant across modes with verbal representation limited to visual matters and are therefore somewhat restricting. In a short film such as *Congo VR* with a limited number of shots (as is currently typical in CVR), time is a valuable commodity. It is worth remembering the alternative in framed, faster edits where cutaway shots can do much work while a social actor is speaking. Without the ability to 'cut away', Alastair Leithead must alternate between two extremes: of being 'in the moment' as things happen so to speak, and then switching voice dramatically to expound on deeper issues while remaining in the same visual space.

Regarding Mamie's re-contextualizing, the issue (if it to be an issue) is that the modes do not combine to present an immediate cross-modal field for the spectator with meanings instead hanging in the air, rather than anchored to immediate embodied experience. Another issue is in the temporal schism that presents in Mamie's re-contextualising. This occurs where the visuals are enhancing the voice-over causally, as conjunction of *result* where she is verbally recounting historic events and visually signifying aspects of their repercussions. It is seen in shot M#5 where she is seen in a process akin to *leaving*, while also recounting her childhood decision to refuse 'FGM' (female genital mutilation). A natural causal relationship emerges where the FGM refusal has led her to this point. Here, the visual take on an ambiguous performative effect that is seen to greater and lesser degrees through the text. It can be argued that the spectator's reflexive sense of embodied time in CVR contributes to a sense of contemporaneity that necessitate logical conjunctions like this. If she is to be experienced in the same 'time' as the spectator, she must, to some degree be showing the repercussions of her verbalised, past experiences. This equates to meanings such as: I refused FGM – and here I am, leaving',[20] or more realistically from the spectator's perspective: I refused FGM – and here I am, showing you *what it looked like* for me to leave. In this regard, the aspect of presence is a limiting factor where scenes such as these become highly performative.

5.4   Concluding Remarks

This chapter addressed aspects of the three research questions. Regarding the first question, CVR is described as a semiotic technology for the ways discourse is configured by the 360° camera and rotational display. My intention when approaching the study from this point of view was to avoid the term 'interaction', as it is a loaded and problematic. *Simultaneity, multiplicity,* and *rotation* proved to be better suited to discussing CVR as they are immutable co-extensive aspects of the form. This chapter described *simultaneity* as the omnidirectional, heterogeneous field of activity.

---

[20] A more direct co-contextualizing 'fit' would be between the image in M#5 (at the doorway with her bags) and the verbal 'threw me and my belongings out on to the street' in M#7.

Rotation was discussed tangentially for the ways it is implicit in the realisation of experiential and logical meanings. Participants, processes, and circumstances were found to be fluid entities in the 360° image. Figures and episodes interact across processes, creating a matrix-like field of activity where Kress and van Leeuwen's *conversion* becomes a feature of the form, rather than a technical aside. The relatively indeterminate nature of who or what is viewed and in which order, means that the order in which transitivity is processed will vary from spectator to spectator, with slightly different meanings. The rotating spectator therefore has access to a mode of low-level meaning making different to the reception of sequential framed representations, and more closely resembling the kind of logical associating one engages in outside of the headset. The inferences made for example when Mamie in M#18 is viewed first as a 'leader of her flock', and then transformed in a quick rotation to a 'newcomer amongst adults', are reflexively the spectator's and the product of a rotation.

In this regard, the rotating spectator can be considered to have a degree of semiotic autonomy when meanings are rotationally determined. Earlier transparent forms such as direct cinema made claims on the technology's ability to present ostensibly unmediated images, where the spectator was given interpretive agency, unaided by the didactic voice. Interpretive agency in CVR can be considered partly as the rotating reactive, sense-making affordances where meanings already encoded in the texts' pixels will be imbued with a degree of spontaneous reconstruction in their reception. Mamie is 'available' semantically in two participant states simultaneously, until the connection is made, and significantly made by the spectator in their own time. I will discuss this further in chapter 8, through the register variable MODE, where the rotating-as-sensemaking will be framed as a fundamental technological characteristic of the form. Circumstance is also transformed in CVR where transactional participants and their processes were found to leave their circumstantial trace as lingering enhancements on other participants and their processes. Essentially, even in the experiential analysis, there was a logical aspect that pervades the 360° image, with enhancements imminent in every rotation.

Where the logical function was addressed directly, the image presented relations of contrast that at times necessitate rotation to resolve. Simultaneous contrast was a suitable means of making conjunctions that had the potential to connect parallel episodes. The BBC team seemed to be at risk in certain shots by virtue of their difference to their subjects, at times presenting odd dichotomies, such as in C#15 where they are seated, drinking, and chatting while the sapeurs enact their dance routine. The logical function is of limited use in singular shots where similarity and difference are of little semantic use without being incorporated into larger meaning complexes. I will provide a macro analysis of the texts in chapter 7 incorporating these simple conjunctions into entire phases of text, where they become useful as logical operators across

modes and shots.

Where the author of the text is concerned, the multiplicity and fluidity of ideational meanings is something that must be reckoned with. I have demonstrated the kinds of meanings at risk in CVR, which point to the difficulty in presenting a singular representation. This is compounded by the apparatus, which requires that the filmmaker vacate the shot prior to recording. A prominent theme in this regard was the ubiquity of circumstance. Visual circumstance is practically impossible to erase in the 360° image and it is as such always at risk of signification. This impacts the production of nonfiction CVR, where there is an ethical and a practical component to stage managing the situations of filming, where participants need to be singled out and others left in circumstantial arrangements. Kool (2016b) describes this as part of CVR's potential for editorialising and propaganda). This was found to be more of an issue in *Congo VR* where certain figures, likely intended to be circumstantial as accompaniment, were visibly interacting in process-vectors with others who had ratified participant status. Junior's accompaniment in C#13 is a case in point, where the woman close-by seems to have importance, but this importance is not realised. The converse was found in *Mamie's Dream*, where the young girl in M#16 is positioned poorly, leaving her in circumstantial 'limbo' as the shot plays out. This is largely a problem of salience and cohesion, and I will elaborate on both textual instances along these lines, in their respective sections 7.2, and 7.4 in chapter 7.

Examining ideation in the texts cross-modally, gave an insight into the choices made in post-production regarding the producers' construction of multimodal meanings, and the kinds of interpretive space allowed. By using contextualisation frameworks, the texts showed a sharp divergence in the ways meaning is formed between speech and image. The texts differ in the ways that language either consolidates local meanings as a multimodal experience of the 'here and now', or produces meanings akin to Cheong's 'semantic effervescence' (Cheong, 2004) that are diffuse and inference based. *Congo VR* takes both paths, signalling its need to both create meanings that are contemporaneous to the spectator, and present rhetorical devices required for the exposition-based subjects.

Regarding the study's contribution to SFMDA methodologies, I presented several ideas in this chapter regarding the VR camera's configuring of ideational meanings. I operationalised a rank system for CVR, where the shot can be considered as the simultaneous field of activity, the field-of-view as the portion of the image currently selected, and the episodes and figures as groups and processes, and their individual constituents. The rank system was fit for purpose for the most part but relatively difficult to operationalise where field-of-view was involved. For example, CRUs took the form of figures and episodes, such as the sapeurs and BBC team in C#15 which

were realised by the rotation of field-of-view but can be related to each other without field-of-view being involved. The difference is that whereas shot, episode and figure denote object hierarchies in the content of ideational field, field-of-view denotes the 'technical apparatus' aspect of rank. field-of-view is also a more indeterminate and fleeting phenomenon compared with the other ranks which are perceptually fixed in the 360° image. As such, field-of-view can be considered an aspect of rank but difficult to apportion any semantic properties in itself.

6. Interpersonal Relations

This chapter addresses the first research question in a similar fashion to the previous chapter, by characterising CVR as a semiotic technology with uniquely configured semantics. Where the previous chapter articulated this through the representational aspects of the texts, this chapter describes matters through the interpersonal function where discourse constructs the spectator.

I will address the second question by discussing the interpersonal ramifications for nonfiction CVR producers. The overarching phenomena of simultaneity, multiplicity, and rotation are described here as a matrix of relative interpersonal associations, where the spectator cannot form a relationship with one entity without it being relative to another. This presents novel ways of using the 360° modality to engender dynamic filmic perspectives, and points to ethical issues regarding identification and the potential for othering of social actors and groups.

The third question regarding the study's contribution to SFMDA is addressed by the specific application and modification of interpersonal resources in CVR. A key integration of Adam Kendon's socio-spatial theories to CVR analysis. These, to my knowledge have not featured in CVR studies to date and are used to add descriptive clarity to the specific interpersonal semantics of the 360° image. I will begin with a description of Kendon's frameworks, as they were applied to the texts and then use them in the subsequent discussions of the interpersonal metafunction.

6.1   Adam Kendon's Socio-spatial Formations

Kendon's socio-spatial theories were described in chapter 3, section 3.5.1 as a means of structuring interpersonal systems into the 360° space as a series of *transactional segments* and *formations*. It bears repeating that these are novel and apt to CVR analysis as they are perceivable in the 360° through the related aspects of rotation, place-illusion, and object permanence. While hypothetically these could also be illustrated through and edited sequence of framed shots, they would not be reflexively resolved and experienced as proximate by the viewer.

A *transactional segment* (TS) is a portion of the overall space delimited dynamically by the transactional requirements ('use-space') of an actor to the necessary exclusion of all else for it to support the activity. *Formations* are more socially determined spatial configurations, where for example *F-formations* coalesce around a central 'O-space' where ratified participants take a position inside, excluding those outside. Examples of TSs and formations are provided in this subsection to demonstrate the phenomenon. Beginning with transactional segments and use-space, I will describe the different ways that TSs manifest in the shots and crucially, how the

spectator is positioned relative to them.

6.1.1   Transactional segments

When the spectator is immersed in a social setting involving a participant enacting a transactional process, they can be attributed a defined position relative to a transactional segment. Having analysed the texts, three basic TS configurations were found:

1. The spectator *external* to established transactional segment
2. The spectator *implicated* in transactional segment
3. The spectator in *violation* of other participants' transactional segment

In the first configuration, the spectator is positioned external to an established use-space that is required for a participant to enact their process. Of the three types, this one results in the most transparent experience available. In the following examples, the spectator, while inhabiting the filmed subjects' space in a general sense, doesn't overlap with a participant's use-space. By not violating others' transactional segments, the spectator is not perceptually impinging on and preventing a participant from realizing their transitive goal. This is extended to their not inhabiting a gaze between participants. *Mamie's Dream* exhibits this phenomenon in M#7 (with her daughter) and M#17 (with the female pupil). Both shots contain clear transactional vectors from which the spectator remains excluded for the duration of the shot. *Congo VR*, as a more reflexive text has fewer examples of this kind, where many of the processes are direct address and where the spectator is engendered to feel part of the mise-en-scene. Exceptions include C#25 where Osambia-Kpwata Fyfy is angled away from the spectator and her gaze is blocked by her sunshades. As is evidenced in these shots, by remaining on the outside of the transactional segment, the spectator can view the transaction *from without,* and as such this configuration is another way of describing a relative *bystander perspective* and contact of *offer*.

The second configuration involves the spectator being *implicated in transactional segments* and reflects a kind of 'transitive-spectator'. Transactions necessarily involve the use of space between actor-goal; speaker-listener and so on and the 'use-space is also the spectator's as they are implicated in a visible process. This is found in both texts in two significant ways, where there is verbal and/or visual address to the spectator, and where the spectator is characterised visually in the mise-en-scene. There is much variance in how these occur and how they overlap with each other, which will be discussed as borderline cases at the end of this subsection. In the first instance, the spectator assumes the imagined role of reciprocator of a speaker's gaze and

hearer of their diegetic verbiage. The reciprocal nature of this arrangement results in a bi-directional set of overlapping segments that shuts out all other exterior participants for the duration of the speech act. With some exceptions, this is found in shots where there is a verbal address to camera. In Congo #21, the pygmy elder sets up a very strong TS with the spectator through his verbal, visual and kinetic gesturing. Considering the mutually exclusive nature of TSs, this shot is of particular interest when the other tribesmen are factored in. As outside of the spectator-pygmy elder TS, they are in a subtle way 'off-limits' where our focus must by convention remain with the elder pygmy. Breaking away from his gaze is a transgressive act, made doubly transgressive when the spectator's gaze encounters the downcast tribesmen in their moment of painful reflection.

Direct address, as described here will almost always involve visual address where a purely gaze-based TS is established. This kind of TS is not necessarily on the same level of magnitude as the first kind however as gazes can be of many kinds, from the focused gaze accompanying speech, to the brief surreptitious glances to camera found throughout the texts (e.g., the cooking woman in M#9 (3:34)). Also, gaze cannot always be considered as use-space, enacted to the exclusion of other aspects of the space. For example, in C#19 the spectator is external to the TS established by the Pygmy man engaged in digging process. The shot also includes a glance by the standing Pygmy man to the spectator, establishing a brief TS enveloping the spectator's position. This suggests that the shot inhabits two simultaneous TSs, one where the spectator is external, and one where they are implicated.

Other ways that the spectator is implicated in a TS is where they are personified through their placement relative to elements of the mise-en-scene. Alastair Leithead's gaze in C#8 and its TS also coincides with a larger fixed use-space where there is a larger TS already in place as determined by the seating arrangement and bottle. Here, point of view merges with the TS where the spectator is reflexively aware of their positioning in an ongoing or immanent TS as they are implicated in a social encounter with Alastair Leithead. C#30 also places the spectator in a TS, standing in the military vehicle but unlike the previous example with Alastair Leithead, the spectator shares an assumed singular TS with the soldier, looking out on to the road.

The third way that the spectator interacts with visible TS's is more transgressive in nature and compared to the other types, is the most hypermediated. This interaction points to the limitations of CVR as well as its ability to use a kind of 'erroneous embodiment' as a semantic device. In this case, the spectator *violates other participants' transactional segment,* and are less of a constructed transactional participant, and more of an incidental intruder into others' use-space. Bearing in

mind that TSs are established to enact and maintain processes, it stands to reason that the perception of their violation, albeit slight, will involve a notional impingement and compromising of that process. Examples of this are found throughout *Mamie's Dream*, such as in M#6 where the spectator is positioned liminally between the teacher and the students, who form a TS dedicated to their observing his movements, an effect replicated in M#11 where the spectator is positioned between Mamie and the blackboard, the object of her focused gaze. In both cases, the spectator cannot maintain an objectivity in the scene as they are in an activated zone. This is compounded by the fact that they must rotate 180° in both directions to resolve the overlapping TSs, adding a somewhat frenetic sensorimotor aspect to the encounter.

M#3 presents an example of a violation where the spectator is positioned awkwardly in a series of low-level overlapping 'footballing' TSs while also inhabiting a larger TS enacted by those watching the footballers from above. This shot is complex from a use-space perspective as it is arguably the spectator's use-space that is violated by the highly kinetic footballers in which case it is a matter of proxemics, where *personal space* is violated. In this shot, it is also the transactional segment established by the onlooking pupils that incorporates the spectator, resulting in the spectator's envelopment in both the footballers and the onlookers overlapping TSs. The spectator is as such potentially violating two overlapping TSs. Other similar examples of complex TS violations can be found in *Congo VR*. C#17 involves the camera-spectator positioned adjacent to the dancing tribesmen while also in multiple gaze-based TSs established by those watching the performance. This is perhaps less of a violation in that the tribesmen's actions, having no vectors, are behavioural. C#12 (2:30) presents a similar momentary violation as a woman walks to the camera at approximately intimate distance and calls loudly to someone further along the platform. In this case the TS is culturally conditioned to allow brief violations.

6.1.2 Formations

Formations were found across both texts, wherever people arrange themselves in defined spatial configurations to enact joint processes: greeting, conversing, spectating and so on. I will discuss formations in the texts, as structured by the physical and institutionalised aspects of mise-en-scene (such as train carriages, boats as seen in the texts), and as purely social and internally 'self-forming' ways of structuring the social space within and around them. Discussion will include the effect of F- formations between filmed participants - and between filmed participants and the spectator.

Obvious formations occur where participants and the camera are placed in a physically

constrained space, that then becomes social. Both films, for example place the spectator in a boat (M#1 and C#4) but with different kinds of formations therein. Congo #4 places the spectator in a front facing seated formation, facing Alastair Leithead who creates an imbalanced formation with himself as leader to the team reflected in the teacher-student dynamic (Kendon, 2010). In Mamie's boating shot, she and the spectator are positioned in a vis-à-vis formation denoting parity of esteem (ibid.). C#8 presents another seated formation with Alastair Leithead and the spectator arranged around a table as part of the wider formation of fellow travellers arranged throughout the carriage. A looser kind of F-formation is found in C#7 where the spectator is positioned on the train platform with others waiting for the train to disembark. The platform arranges the formation of mostly side-by-side and facing the train. Naturally, the spectator can transgress this formation and face any direction they choose; what is important is that the formation is available for them to relate to, should they so choose.

Certain formations do not require spatial constraints and are enacted by the spontaneous needs of those involved. 'Vis-a-vis formations' were found throughout *Congo VR* where there is a diegetic speaker acknowledging the spectator through direct address. Alastair Leithead, Junior, Princess Mamicho, and the Pygmy elder all enact vis-à-vis formations. Junior also highlights a potential for formations to be formed erroneously, complicating the shot and the nature of the relationships of those involved. An example is found in C#13 which arranges the spectator in a complicated formational arrangement relative to Junior and the woman. The formation is either an F-formation involving all three or a vis-à-vis that the spectator is adjacent to (Figure 14, below). Regardless, once the shot commences Junior immediately draws the spectator into a vis-à-vis (3:04) with him, through his direct address. The formation-switching presents a socio-spatial 'jump-cut' of sorts, drawing attention to the artificiality of the exchange. Formations are useful in this regard as a descriptive device for the issues that CVR poses in its lack of framing and simultaneity.

F-formation including Junior, woman and spectator   Vis-a-vis formation including Junior and spectator
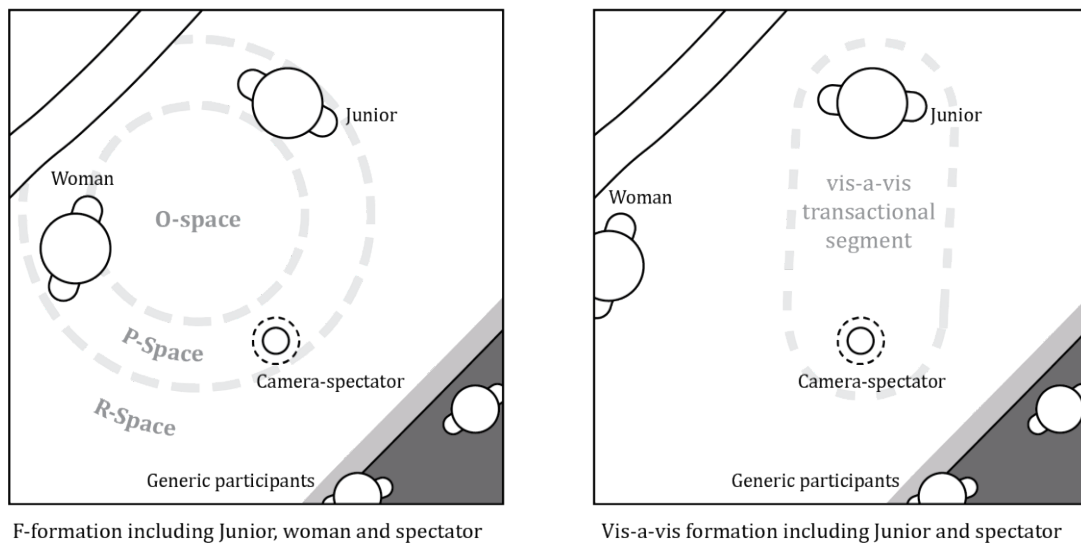
Figure 14: Mutually exclusive socio-spatial arrangements in C#13

Formations take on a particular significance when entities (filmed or spectator) are found to be placed outside relative to a formation, signifying social exclusion. Figure 15, below shows a prototypical F-formation occurring in the busy Kinshasa train station where the spectator is positioned at a personal distance to a group of travellers who have congregated in an inward-facing circle, greeting each other having disembarked from the train. This is a prototypical F-formation as its inhabitants have 'closed ranks' to maintain a protected o-space from the bustle of the surrounding crowds. The spectator is positioned immediately adjacent to the formation in r-space. This is a charged zone and generally avoided by passers-by and only traversed if a formation member opens the 'p-space' (the circle of participants) to ratify and allow an outsider in. The effect is naturally mitigated by the specific social situation, where it is accepted to be momentarily at close quarters to fellow travellers but is nonetheless a clear example of a formation involving (by excluding) the spectator.

Figure 15: Prototypical F-formation in Congo VR (2:27)

Mamie is positioned outside of a formation in two shots with different effects. In M#8, she is seated above and behind a group of market sellers, who are arranged in an L-shaped dyadic F-formation, conducive to non-personal conversation (Kendon, 2010). She is too far away to be in violation of their formation, but her isolation is nonetheless compounded by these people being in a notable interaction with each other contrastive with her being alone. Mamie is again on the outside of a formation in M#13 but in a manner that arguably complicates the preferred reading (for this reason, I will use the voice-track as a means of supporting my claim). When this shot is considered cross-modally with image, voice, and music, the overall narrative is of Mamie's burgeoning status within the group ("But now I am a community leader and a role model"). In the shot, she moves through a busy social area comprising multiple formations of people engaged in various activities. Some formations are structured around common activities which neither involve nor exclude Mamie and are secondary in terms of their semantics (watching a man performing 'keepie-uppie' with a football, for example). Mamie's 'walkway' is empty and unobstructed (for practical reasons relating to the moving camera) and the closest cluster of people, both seated and standing create a loose F-formation. It is not as tightly organised as the example given in *Congo VR* (here, a single participant stands still facing outwards) but nonetheless has a defined border and o-space between a handful of the participants. Mamie passes by the formation and gestures a greeting that is visually unreciprocated. Her greeting is not unnatural, she may indeed know one or a few of the participants. The lack of reciprocation, when paired with the incongruity with one's usual behaviour while in o-space (ignoring the formation for fear of drawing attention to one's unratified status) places her at a perceptual social distance counter to the meanings in-play in the voice-over.

122

6.2   Point of View and Perspective: 'Who Is the Spectator?'

Point of view (POV) was examined across the texts for how the spectator is characterised as having a subjective position in the field of activity, in other words 'who's eyes are we seeing things through'. Using Branigan's nested subjectivities as a guiding frame, I examined the texts and found two overarching kinds of POV: those that assume a human personage and those that do not. Both texts alternated their use of POV ranging from shots intended to make the spectator feel as if they are an active acknowledged participant, to shots that appeared to abandon this endeavour in favour of more conventional shot types where the CVR camera was used simply in place of a framed camera. This can be described as two ends of a continuum with 'personified POV', at one end where the camera is a character, to 'technical POV' where the camera is not intended to show the spectator's subjective view. Where POV attributes an *identity* to the spectator's view (albeit a notional one), *perspective* describes the nature of the engagement with that which is viewed, as being more, or less, subjectively involved. Analysis involved looking for conventional objective/subjective shots and where they became complicated by the form. Taken as a whole, the analysis provides an umbrella concept in understanding CVR and a useful frame for other interpersonal systems such as *contact* and *power*. Analysing POV and perspective also necessarily involves use-space which will be used here to add descriptive clarity to the analysis.

Both texts constructed a personified spectator, resembling the traditional first-person 'subjective shot' in framed filmmaking, where the viewer is consciously adopting the view of a character in the scene, through who's eyes we see the film. The distinction here is that in framed film there is typically a character already established and where a reverse-shot edit is used to prompt us to go from 'outside' to 'inside' the character's POV.[21] Some VR texts simulate an explicit embodied POV 'in a character', through various post-production techniques (for examples and discussion, see Tong, Lindeman and Regenbrecht, 2021) but this is not typical in CVR and especially nonfiction CVR where the spectator is personified reflexively, essentially as *themselves*. In these texts, there is no 'character' for the spectator to inhabit but there are subtle cues given to imply the presence of the spectator. As such, the spectator's identity remains in a grey area regarding POV, where they are reflexively themselves, yet different 'versions' of themselves.

---

[21] Examples exist of films comprised entirely of first person POV but these are exceptional and rare.

6.2.1   Personified POVs and Subjective Perspectives

The highest degree of personification was found where there was explicit acknowledgement of the spectator ('direct address') combined with elements of the mise-en-scene suggesting the spectator's current, immanent, or past engagement in a transactional process. C#4 presents this configuration, where Alastair Leithead is speaking directly to the camera-spectator, who is 'seated' with the other crew members experiencing transitivity kinetically also through the boat's movement. This is approximated later in C#8 where he and the spectator are part of the train- carriage arrangement, but here his acknowledgement of the spectator is enacted by gaze only (a 'mild' form of direct address). Here the mise-en-scene is used more consciously to construct a 'co- traveller' spectator replete with their own seat and water bottle. The use of a prop here suggests an attempt to merge experiential and interpersonal resources, and to make an implicit POV explicit. To a lesser degree, M#1 presents another boat inhabitant POV, suggesting similar transitivity of *co-travel* also but where Mamie does not acknowledge the spectator at all, lessening the effect of personification and making the POV more generalised.

6.2.2   Bystander POV and Relative Objectivity

It has been argued that CVR's immersive and reflexive nature makes it an inherently first-person POV, thus incapable of presenting objective perspectives (Laurel in McRoberts, 2017; Raphaël in Tricart, 2017, p.99) and there is credence to this claim when CVR is compared to other filmic forms. It is more fruitful to abandon 'objectivity' as a concept in keeping with Branigan's ideas and instead look for relative degrees of subjectivity. Perspective inhabits a continuum, where on one end the texts present configurations that skew towards neutral views, where the information available to the spectator appears to unfold independently, and where the spectator's presence is 'passivized'. These views are largely dependent on the array of processes involved and the spectator's interactions with unfolding transactional segments. In M#12, the spectator is positioned at standing height at the blackboard's edge with Mamie and the pupils forming a bidirectional overlapping transactional segment to which the spectator is tangential. While Mamie does move towards the spectator as she writes on the board, the board itself delimits her activities, which will not spill out into the spectator's immediate space. The sum of this is that the spectator is afforded a relatively passive perspective on events. Taken in context, this neutrality is relative: relative to the overall sense of presence experienced in the shot, and relative to the highly subjective shot that preceded it, where the spectator is inserted into a series of lively transactional segments.

6.2.3 Mixed Perspectives

Many shots provide grounds for considering *mixed,* ambiguous perspectives within a single shot. A phenomenon much like Kress and van Leeuwen's *conversion* emerged in places where the spectator's degrees of subjectivity fluctuate depending on whether the spectator feels themselves to be the subject or object (or both) in the multitude of interactions unfolding around them. In this regard, Branigan's conception of POV as a hierarchy of subjectivities is realised in a stark and novel fashion in CVR. C#10 is a case in point where the spectator is engulfed in a complex matrix of perspectives as subject and object dependent on their current field-of-view. There are two principal groups in the spectator's immediate vicinity, the queue of travellers making their way past the spectator and an F-formation established by the adults and children who have disembarked the train. Like the example given of M#12 above, the F-formation by its nature is closed and where the spectator is *outside looking in,* with a more objective perspective. The other group, in their movement around the camera-spectator and surreptitious glance to camera (e.g., man in green t-shirt (2:23)), place the spectator *on the inside.* This dynamic is repeated elsewhere such as in C#17 where the pygmy men are object to the spectator's gaze, who is violating the TS of others also watching and thus cast into a simultaneous objective/subjective dichotomy.

An interesting arrangement presents in M#4 which most closely resembles a conventional point of view shot. Importantly, the shot involves a moving camera that tracks in towards Mamie as she alternates her view from (our) right to left. As the spectator arrives close to Mamie at the end of the camera's movement, the spectator experiences *her* alternating gaze as she shifts focus to the girls playing close-by. The natural tendency in this situation is to follow her gaze, meaning that we adopt her point of view momentarily. For this effect to work, there was an initial period where Mamie was contextualised at a distance objectively in the scene prior to the effect of joining her perspective. The overall effect is reminiscent of framed edit sequence incorporating a long-shot, medium, and point-of-view shot. The significant difference in CVR however is that the spectator maintains a degree of their own reflexive point of view rather than being 'given' her view as would be the case in framed continuity editing.

6.2.4 Technical Perspectives

Where the spectator does not have a role that is feasibly supportive of human placement, the POV becomes 'technical'. Both texts exhibit technical POVs in different ways such as the use of aerial shots where the camera is attached to a drone. *Mamie's Dream* has just one such shot (M#19) and

the film's producers chose to erase the drone in post-production. Here the view becomes a generalised view: an *idea* rather than a personified entity's narrated account. The aerial shots in *Congo VR* differ in that the drone remains clearly visible. This has a different effect on personification where the proximity to the drone nullifies a generalised cinematic perspective. Here, it is neither personified, nor is it cinematic but inhabits a grey area in terms of the embodied spectator. This stark effect is repeated in C#11 also, where the camera-spectator is mounted atop a taxi edging its way through Kinshasa traffic. Finally, M#2 presents a technical perspective that is more nuanced. Here, the spectator is attached to the outer edge of the bridge with no visible support below (to 'stand on') resulting in a sheer drop to the river below. There are no drones or vehicles here to block the feasibility of a human personage, but the effect is no less stark as there is no way to resolve a 'human-like' position for the spectator. Technical POV in CVR is difficult to describe using conventional means. Returning to Branigan, if this is not perceivable as the view of a notional 'character', then it is the view of the filmmaker and their subjective viewpoint on events. If the BBC crew are to be interpreted as the producers of the film, then it is perhaps their intermediate view that is visualised *for* the spectator (as opposed to the spectator taking a view). The aerial shots are in this regard *their* attempt to capture the essence of scale and location, and the taxi-mounted camera is used to convey their experience of DRC traffic. Regardless of whether we locate this subjectivity to Alastair Leithead and his fellow team members, or to an implied 'BBC POV', the overarching distinction here is that it is not the spectator's personified subjectivity that is encoded into these shots.
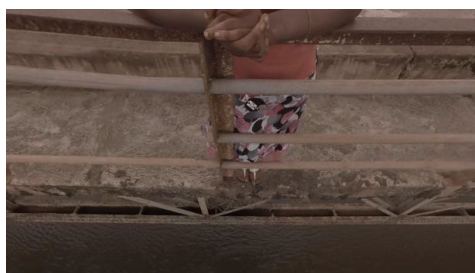
Mamie's Dream, shot 2 involving an unsupported spectator (0:46)

Congo VR, shot 11 involving a non-human spectatorial position (2:40)

Figure 16: Technical perspectives in each of the texts

Examining CVR in this way leads back to the issues that CVR faces as a perspectival medium, and the question of who the spectator is 'psychically merging' with. The technical POV presents a schism that comes back to the lack of a frame and distancing picture plane and the CVR spectator's reflexive, rotational and proprioceptive 'felt' embodiment which cannot be switched off in the

360° space. Where the personified perspective is implied, the spectator can at most be considered to 'merge' with their own view – as if they are experiencing the scene directly. Where the spectator is placed in a technical POV, the dual perspective required involves them identifying with a contorted experience of their own point of view. Branigan's hierarchy of POVs positions the viewer's POV as the ultimate view, and in a sense, this view is collapsed in CVR into the lower-level perspectives involved in visual narration.

6.2.5   Contact

Discussing POV and perspective necessarily involves discussing *contact*: where conceptions of perspective are modified in CVR, as discussed above, there will be a correlating effect on contact. An 'image act' of *offer* in CVR is not as stable a concept as was first conceived in visual grammar, which is essentially a framed, still image construct. Kress and van Leeuwen describe an offer where the viewer is an 'invisible onlooker' presented with "items of information, objects of contemplation, impersonally, as though they were specimens in a display case" (1996, p.119), and *demand* where the subject in the image "acknowledges the viewers explicitly, addressing them with a visual 'you'" (ibid 1996, p.117). The 'display case' metaphor presupposes the image frame and picture plane; CVR, lacking the distancing power of the picture plane makes such definitions problematic. CVR's reflexive nature also means that the immersed spectator cannot take a detached perspective as an 'invisible onlooker' when their presence is implied by other spatial and socio-spatial means. Invisibility is a difficult construct to square with presence, and particularly so when one feels awkwardly situated in another's use-space. Also, demand, considered in visual grammar as a sole by-product of gaze is also problematic. Where the demand gaze in still images was conceived whereby the subject acknowledges the spectator's presence, that acknowledgement is achieved by other means in CVR. Also, if gaze is to be considered a primary determinant of demand, the variance in the kinds of gazes-to-camera found in the CVR texts suggests that it is not always a reliable indicator of contact. I will discuss these issues using examples from the texts which were examined for offer and demand, as typified in visual grammar, and for the ways that CVR challenges readings of contact. I will elaborate regarding contact in CVR as a relative quantity, where although it is a less stable concept in CVR, some shots are clearly more offer than demand, and vice versa. Additionally, contact, much like perspective is described for being essentially fluid, where global 'image acts' give way to the acts of individual figures in the 360° image.

Contact of offer is found in both texts but in terms of demand, they diverge significantly along stylistic lines: *Congo VR* uses direct address, whereas *Mamie's Dream* eschews any

acknowledgement of the spectator. In this regard, *Congo VR* skews largely towards demand and Mamie towards offer, typical of naturalistic formats (Kress and van Leeuwen, 1996). All the direct address shots in *Congo VR* can be considered demands, by virtue of explicitly conferring a 'visual you' onto the spectator. The pygmy elder in C#21 approaches a quintessential demand where gaze and kinetic resources combine to enact a sustained visual demand expressing 'you must know this'. Princess Mamicho makes a similar visual enjoinder for spectatorial engagement. In her case it is her involvement, gaze and kinesis that demand that the spectator appreciate the nature of what she is saying (she touches her clothing, gestures to the spectator, and rotates her body, dancelike at the end of her act). It is the sustained and multimodal nature of these acts that marks them out here as prototypically demand.

Unlike still images, contact in CVR cannot be deemed a demand, solely by virtue of gaze to camera. Shots from both films include less overt and often surreptitious glances towards the camera-spectator. Congolese train travellers as well as Junior's acquaintances all glance briefly to camera: the cooking woman in *Mamie's Dream* (M#9, 3:35) casts a glance to the camera, as does Mamie herself in (M#8, 2:57). These gazes are on a different order of magnitude to Ilungama Ayanda and Princess Mamicho's gaze and are also inherently different in kind to those found in still images (there are no 'brief glances' in still images). Rather than fulsome demands, enjoining the spectator to listen and act, these furtive glances serve to disrupt an otherwise offer by briefly heightening the reflexive presence in the scene. Elizabeth Cowie describes these looks in framed documentary as "a spectacle that confronts the spectator in her or his own looking' (Cowie, 2011, p.12).

Offer and demand in CVR fluctuate in the same manner as perspective and POV. Demand, as inherently intersubjective overlaps conceptually with the personified spectator, who is both subject and object within a defined transactional segment. Using the bystander construct discussed above, where the spectator is outside of a TS and not required to rotate their view, the information presented to them can be viewed with a sense of detachment. This was described in M#12, where her teaching activities can be considered as predominantly offer-based, as her and the pupils use-space is established in a single field-of-view for the spectator to contemplate without their constructed position in the scene being impinged via gaze or the need to rotate their view. Furthermore, if CVR's simultaneous multiplicity of perspectives is considered a contributing factor in contact, offer and demand can co-exist across components of a single CVR image. C#10 was used as an example where the spectator is both 'outside' an F-formation who are objects to their gaze, while also themselves the object of the queue of travellers, as gaze (albeit brief, (2.23-2.24)) and kinesis (the woman moving around the camera's position to progress along the platform). Perhaps more apt is C#21, described above for the pygmy elder's forceful demand,

which involves multiple participants who are not acknowledging the spectator and can be described as 'objects of contemplation' rather than actively demanding the attention of the spectator.

6.3 Power, Status and Association

The study's second research question is concerned with the implications for the producers of nonfiction texts, and this involves a sub-question: How does the nonfiction CVR text construct the nonfiction spectator? This subsection addresses these questions directly by describing the effects of the of camera's positioning in the construction of the spectator. Social distance and vertical status are discussed for their respective contributions to the spectator's perceived role in the filmed setting. They are also discussed as factors contributing to a larger complex of *associations* in the 360° image where distance (as horizontal physical distance) involves the spectator being both *close to X, -* while *closer/further to Y*. Vertically, the height of the camera will produce similar effects in terms of being *as tall as X – and taller/smaller than Y.* In both instances, the spectator will be aligning themselves with certain figures and episodes, while being contrastive with others. To discuss these phenomena, I will elaborate social distance and height variations in the texts for how they instantiate singular relations between spectator and filmed subjects, and how they are constructed by their status and associations amongst all others in the wider matrix. In doing so I will point to CVR's semio-spatial novelty and some fundamental ethical constraints of the 360° image.

6.3.1 Relative Social distance and involvement

I will briefly reiterate how CVR differs from framed imagery regarding social distance and put some necessary caveats in place. Framed images' effect of social distance is generally conceptualised as the 'size of the frame' with frame-based connotations for pseudo-social bonds (Kress and van Leeuwen, 1996). I will not focus on typical connotations of social distance, such as close-up shots and their emotional effects, as these are largely moot in CVR (cf. Dooley, 2020). Rather I will focus on the indexical kind of embodied closeness that CVR facilitates, quantifiable as proxemic distances. Essentially, distance is not considered to engender the same effects of extreme emotional closeness available in framed filming but can provide cues to degrees of social alignment (intimate, personal, social, and public). The texts are described in this light for the physical distances between singular participants and the angles they take up with each other and the spectator. Non-interpersonal distances are excluded, such as the extreme aerial distances

found in *Congo VR*.[22] These distances are at times far beyond a meaningful intersubjective proxemic (*public* being the maximum at 25ft+). I termed this earlier as a 'technical perspective', which in conventional cinematics is not intended to be read as seen through a subjective narrator, rather the filmmakers themselves who are using it to generalise and 'explain' something of the location (e.g., its vastness and topological interest).

To fully grasp the effects of associative distance in CVR's complex interpersonal scenarios, it is necessary to consider the co-extensive systems such as *involvement* (the relative angles of participants) and *use-space/formations*. Involvement always modulates the effects of proxemic distances, where for example an *intimate* face-to-face distance (six inches) becomes *personal* in a side-by-side arrangement (see Figure 17, below left) and where a *personal* side-by-side distance at eighteen inches becomes *social* when considered in a tandem arrangement (see Figure 17, right).
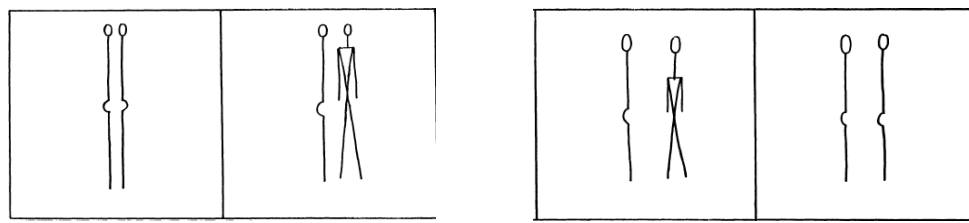
Figure 17: intimate and personal at 6 inches; personal and social at 18 inches (from Martinec, 2001)

Where proxemic distance is used as an interpersonal resource in the texts, *personal* distance was predominant in both texts where, "Subjects of personal interest and involvement can be discussed at this distance. Head size is perceived as normal and details of the other person's features are clearly visible" (Hall, 1966). It is therefore a distance conducive to direct address and/or where a salient figure is enacting a low-level process. In *Congo VR*, seven shots contain specified diegetic speakers and all but one of these shots positioned the spectator at a personal distance to the speaker. Alastair Leithead, Junior, and the Pygmy elder each speak at roughly arm's length and with a high degree of involvement, replicating prototypical vis-à-vis conversational use-space. Princess Mamicho is closer to the spectator, bordering between personal and intimate distance (1.5-2.5ft), engendering a higher degree of familiarity reflected in the other modes. Osambia- Kpwata Fyfy on the other hand (C#25) is positioned further than personal distance, at a *social* distance to the spectator (3-12ft), widening the gap physically and socially to a distance befitting her self-proclaimed status as 'traditional chief' and 'Mobutu's

---

[22] Congo VR's aerial views are found in shots 1,2,3,5,12,16,23,29.

daughter'. An outlier in *Mamie's Dream* is found in M#5 where Mamie is positioned bordering on a social-public distance. Here, distance is used to allow her to be perceived as part of an imminent transactional process of some sort, breaking the pattern of close and conceptual arrangements.

Where the distance is altered in-shot through camera movement, social distancing becomes a dynamic and more visceral interpersonal effect. M#19 uses an aerial shot but unlike *Congo VR*'s, it is significantly lower allowing Mamie to be a fulcrum that the spectator is progressively brought closer to, and then away from. In one sense, this kind of shot is typical and found in framed films where, because of a distancing camera movement, the subject becomes progressively smaller in frame, merging with their surrounds. This is achieved in this shot, with interesting effects on her symbolic attributes relative to her surrounds. Here however, the indexical position and movement differs. Without a frame, the movement is simultaneously front-facing as the spectator perceives themselves to be *leaving Mamie* and also *going-somewhere*. M#4 presents another moving camera in M#19 but at much closer distance. In this shot, Mamie is at a social-personal distance and through a steady repositioning of the camera (akin to a 'dolly' shot) the distance becomes personal. Here the re-distancing achieves a move from impersonal to personal, with the reverse of the effect found the previous example (M#19). Here the camera is 'subjectifying' Mamie through re-distancing, facilitating *our* perceiving of *her* look and its object (the girls playing (1:33)).

The common factor in the examples given thus far is that they have a similar kind of involvement, in other words, they are all facing the spectator. Image analyses seldom incorporate side-by-side arrangements[23] as it is counterintuitive to expect a viewer to take a position 'besides' the photographed or filmed subject. CVR, as rotational and supportive of a heightened object permanence does allow the perception of this arrangement. This will become a factor in the discussion of multiple social distances in a singular shot, where not all participants will be oriented towards the camera.

When social distance is considered in the context of the 360° space, distance as described above, will be augmented in the wider matrix of multiple distances. The texts were analysed for instances where the spectator is close to one participant in the CVR scene, and either *closer, as close*, or *further from,* another. As would be expected, this tends to be rather complex when all

---

[23] *Involvement* in itself will include a filmed subject in a side-on arrangement to the camera, but the camera (and viewer) will nonetheless face toward them. As such this will not be a side-by-side arrangement.

participants are included. To manage this complexity, I will discuss distance shots for the spectator having either *parity of association* where they are as close to all significant participants, or *contrasts of association* where the spectator's positioning suggests they are relatively more aligned with one, than the other. The overarching differences are expressed as a system network in Figure 18, below.
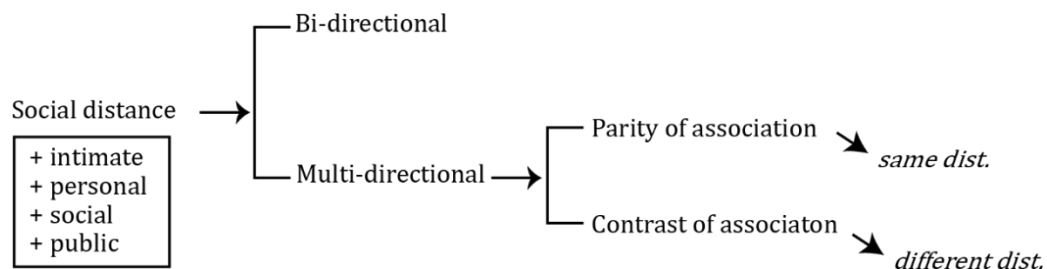
Figure 18: System network illustrating choices in social distance

Parity of association is found to contribute semantically to *Mamie's Dream* where she is seen to be elevating her professional and social status. In M#9, Mamie remains at the typical personal distance to the spectator but is now contextualized amidst other participants who are also at personal distances to the spectator. While the distance to Mamie has not changed, the multi-directional equidistance with Mamie, the readers, and the cooking woman, has an equalizing effect on Mamie. A similar effect is found in M#11 where the distance from spectator to Mamie is roughly equal to the distance between spectator and teacher. The effect here is of engendering in the spectator a parity of association with both Mamie and the teacher, elevating her status in the process through the inferred equality with the teacher.

Where participants are not of equal standing in the text, parity of association can be a complicating factor, placing participant status is at risk. When proximity is used to denote parity of esteem, *generic* participants who are granted such parity with *specified* participants, can present ambiguously. C#13 can be discussed in this light, where the unnamed woman stands with Junior as he speaks and remains roughly equidistant to the spectator throughout the shot. Parity of social distance (as well as the formation issue discussed above) contributes to the temporary elevation of her status, which is not realised or referenced elsewhere in the text.

Where distances are not the same, contrasts of association emerge. As an overt example, C#30 'attaches' the spectator to the military vehicle, at a personal distance to the soldier (intimate plus side-by-side: as per Martinec's schema). In doing so, the spectator is simultaneously at varying public distances to those sharing the road. Other less obvious examples include M#18, which was

132

described previously for transformative transitivity involving Mamie, dancing with the pupils, watched by the seated adults. The interpersonal dimension to this dynamic can be also described partly via their relative distances to the spectator. Mamie is at a public distance to the seated group and the onlooking woman. The camera, positioned at a personal distance to Mamie sets up a relationship of contrast as we are *with* Mamie and *'not with'* the others. The association is clear as the relative distances are not used in the same fashion as in M#11 (where she is on equal standing with the teacher). Here, Mamie can be read as subtly isolated, a reading that perhaps complicates but does not negate the overall signification of the scene where she is to be read as 'being her own person', a reading supported by the following shot at the river.

Use-space adds qualitative information to these associations, where the effects of the spectator's relative distance will differ depending on whether they are positioned within, or outside transactional segments. C#15 Illustrates the combination of these resources. Here, the spectator is positioned closer to the dancers (personal-social) than the onlooking seated BBC crew (social) - but closer to the BBC camera operator crouched below (personal). The relative distance from spectator to camera operator (an extension of the BBC crew) sets up a close, stable association with him. In this instance however, 'personal' (at approx. 2.5 feet) according to Martinec is read as *social* and as such should be further socially distanced. Here however, use-space becomes an augmenting factor: when the spectator is rotated towards the dancing sapeurs, the spectator and camera operator are close enough to be enacting a joint *spectating* TS (a side-by-side spectating formation (Kendon, 2010, p.10)), directed at the sapeurs. In this instance, associative distances, involvement and use-space can be described as combining to socially construct the spectator with a POV in the space.

6.3.2   *Perspective: power and association*

In conventional cinematography, the vertical positioning of the framed camera plays a part in determining how filmed subjects appear, as having more or less power than the camera/viewer. Analysis of the texts showed that this is used in CVR also where both texts instantiated power relations between spectator and subject by their relative verticality in the shot. Certain constraining factors must be reiterated briefly as caveats for the following discussion. First, the CVR camera, lacking a frame, has no directionality and cannot present the low and high tilted 'angles', generally used in framed filming to denote relative power between viewer and subject. Rather, the spectator will have a *height* from which they can choose (or choose not) to tilt their field-of-view to achieve similar effects. As such, CVR only affords part of the overall effect, and the spectator is invited to fill in the rest. Other complicating factors when translating this filmic

effect from conventional cinema to CVR relate to the sense of embodiment that is intricately related to the perception of height. In CVR the height of the camera will translate to an embodied height, felt by the spectator, who may feel *tall*, *short*, *standing* or *sitting*. Added to this, extremely low or high vantage points will distort the perception of the filmed space, potentially disorienting the spectator (Passmore et al., 2017; Rothe et al., 2018). As such the range of heights available to the CVR producer is significantly restricted and will be reflexively held to index embodied information additional to any filmic effects regarding power.

The aspect of embodiment also opens the analysis into novel territory where, much like social distance described above, there will be a matrix of multiple height differences with the spectator positioned above, below or at shoulder-level with filmed subjects.[24] In this regard, the shot opens analytically to a wider *distribution* of statuses and associations. For the reasons outlined above, there will also be a strong overlap with POV where the spectator is *standing with,* or *seated with* participants. As with social distance, I will begin the discussion with one-to-one height relationships, as power differentials between the spectator and a single participant. I will follow this with descriptions of verticality in the context of the wider associations, discussing the effects of verticality and the issues it raises for CVR.

Shots that place the spectator below the subjects' shoulder-line are used effectively in the texts as a way of supporting the dominance of those filmed, as they verbalise their statuses and power. In M#15, Mamie and her pupil are tracked by a moving spectator positioned lower than the pair reinforcing Mamie's verbalised status, "My success is also important for the young girls that I teach". Similarly, C#4 positions the spectator lower than Alastair Leithead as he verbalises his identifying process "I'm Alastair Leithead the BBC's Africa correspondent" followed by his agentive, actor-goal process "And the team and I will be taking you on a great adventure up this magnificent River". Other less clear examples include M#2, where the spectator is positioned at roughly Mamie's shoulder-height as she stands on the bridge. Here the spectator is at an intimate distance however and this proximity forces an upward gaze to view her face, positioning Mamie in a dominant status relation (compounded by a vulnerable, precariously placed spectator). This gives Mamie an expansive demeanour that is possibly at odds with that portion of the film ("*My father said,* "Girls, they just get married and move away""). Finally, M#3 presents a notable instance of the spectator being at significantly lower vantage to multiple participants, lending an exposed character to the height difference where a full row of pupils are themselves

---

[24] Shoulder-level-shots are thus named as they position the camera at approximately shoulder height, to give the impression of being eye-to-eye without presenting an awkwardly composed image with a large empty space above the subject's head.

'spectating' the unfolding action at ground level, including the spectator.

The spectator was found in the texts to be positioned above a filmed subject, but with more varying kinds of effects than where they are below. *Contact* has a part in determining the nature of this dynamic, where *demand* images presented the most striking power difference. C#21 places the spectator higher than Ilungama Ayanda who speaks directly to the camera. This is prototypical as the upward gaze-vector implies he is speaking from a position of weakness elaborating the subject matter in the voice-over ("They were forced to harvest these rubber vines for the Belgian masters"). M#8 presents a similar example of visual submissiveness, albeit without direct address (excepting her brief glance to camera (3:07)). Here the spectator is at true eye-level with Mamie making her appear slightly compressed, an effect that is compounded by other resources such as her disengaged gaze, her posture and lack of movement, which combined correlate with the voice ("I was alone").

Where the spectator is above and contact is more of offer, participants tended to be engaged in processes below the spectator's vantage. This does not necessarily denote power as processes that are occurring at ground level requires the camera to be at a relatively high vantage. Congo #19 and 22 are similar in that they each present ground-level processes (digging and fire-lighting) where the spectator is above partly by necessity. A ground level or worm-eye view would distort the sense of embodiment (Passmore et al., 2017) and for the spectator to fall within the range of a 'natural' i.e., personified height, they must be above looking downwards. C#17 presents a similar issue where the camera is above the pygmy men, thus making them appear compressed in the field-of-view. Here the necessity may be to avoid making the *non-pygmy* participants (e.g., BBC crew) look enormous by filming the pygmies on their level. It is also likely a matter of semiotic choice in that the height difference are a way of highlighting their diminutive stature. This raises questions as to how a group such as the pygmies *should* be represented. What is of more interest here however, is the fact that in the absence of framing and editing, they *must* be represented either on their level, while making all else feel giant-like, or filmed from above with implications for how the spectator is expected to relate to them.

Height in CVR gives rise to identification and POV which, in the case of the pygmies can be representationally problematic. I have suggested that the height of the camera is either a choice or necessity but either way the spectator is *above pygmies because pygmies are small.* The issue that arises here relates as much to the height of the camera relative to the other *non-pygmy* participants in the scene. The BBC crew and their adjunct DRC assistant are shown as roughly at the spectator's height forcing the spectator to negotiate two coalescing POVs, as *taller-than*

the pygmies and *same-as* the rest. This construction requires that the spectator to therefore identify by association of heights as 'non-pygmy'. This in turn has a bearing on social distance where to be 'not-pygmy', yet in their personal space presents an incongruity that could be described as a hallmark of CVR.

The texts present a curious phenomenon suggesting a way of reading height differentials reminiscent of conventional framed POV methods. M#6 involves an adult male interacting with a seated group of students. The shot commences with an initial vertical relationship with the spectator above, and pupils below. This is resolvable by their being seated and the spectator assumed to be standing. The shot is problematic however when the teacher enters the room and briefly takes a position where the spectator is 'inserted' into the established use-space between teacher and pupils. The teacher's height is above the spectator, creating a status arrangement with implications for perspective. When the vertical configuration (pupils < spectator < teacher) is considered independent of field-of-view, the spectator is in a sense 'splitting the difference' height-wise (found in C#19 and 22 also). The effects can be described here in the context of a neither-nor status but also as a rare occurrence where the spectator can be attributed with potential perspectives of both filmed subjects. When gaze (and its use-space) and rotation are factored in, a diametric contrast of power emerges on either side of the spectator (pupils-teacher). When the spectator turns to face the class, the spectator looks down (like/as the teacher). When the spectator looks towards the teacher, they look up (like/as the pupils). There is an effect closely resembling a shot-reverse-shot, where each alternating perspective is 'cut' into the shot by the spectator's rotating their view. This effect is problematic and fleeting as the teacher doesn't extend a focused gaze to the room and quickly moves away, breaking the configuration. It is problematic on a more fundamental level through the spectator's own reflexive perspective, an issue discussed previously in a general sense and where there is a violation of the transactional segment between teacher and class, established by his movements their reciprocated gaze.

6.4 Interpersonal Cross-modal Organisation

Metafunctional organisation in the interpersonal differs significantly to the ideational. Where the ideational function permits the cross-modal analysis of open-classes of things (people, their processes, and circumstances), interpersonal language resources are structural and bedded in grammar (e.g., MOOD and modality). There is therefore less freedom in this function to move back-and-forth between verbal and visual components as interpersonal structures differ fundamentally between modes. For example *social distance, involvement* and *power* do not have

direct lexicogrammatic counterparts, and *modality* cannot be compared directly between speech and imagery, a point made by Kress and van Leeuwen (1996). For this reason, I kept analysis at a high level, restricting interpersonal cross-modal organization to correlations ('reinforcements' (Royce, 1998)) between *visual contact* (offer and demand) and *MOOD*. The films in question, being largely exposition based, constrain variations in these reinforcements in their MOOD, which, barring brief exceptions, is in the unmodulated indicative declarative with few questions or requests*.* As such visual-verbal congruences in the films are for the most part either *offer-offer* or *demand-offer* (statements being offerings of information). A singular moment of cross-modal *demand* is found in C#14 (Junior and Princess Mamicho in the yard) where Princess Mamicho tags her statement with an ambiguous interrogative/statement "A sapeur is an artist and into fashion, *you see what I mean*". The main statement is a declarative and the tag functions to modalize her sureness by seeking the spectator's consensus. Her proximity, gaze (albeit behind sunshades) and kinetic gestures form an overt demand that is congruent across the modes. This *demand-dema*nd reinforcement, albeit fleeting is notable as an exception where other configurations make up the bulk of the interpersonal cross-modal constructions.

Offer-offer reinforcements are found where objective lexis and imagery coincide. I used Mamie as an example in M#12 (teaching) where she establishes her use-space exclusive of the spectator, thus supporting a bystander perspective equating to offer. Mamie's teaching is as such congruent with the declarative *offer* of information in the lexicogrammar, 'the students that I teach call me teacher', where the interpersonal function organises the shot as a *multimodal offer*. The aerial shots in Congo also create reinforcements of offer, with perspective removed a degree further across the image and lexicogrammar. C#1 for example has no visible human participants and is as close to a third-person objective perspective as can be achieved (visible drone notwithstanding) with verbal declaratives also concerning third parties ('the Mighty Congo River […] Democratic Republic of Congo'). *Congo VR* repeats this congruence over the ensuing shots albeit with diminishing objectivity (e.g., 'beautiful but troubled') presenting similar aerial 'offers' throughout the film intended to orient the spectator geographically and supply necessary exposition. These examples can be considered prototypical offers, involving a level of objectivity that is less typical in both films and perhaps problematic given CVR's semiotics more broadly.

Potential issues arise where the overall construct is exposition, but the complicated nature of offer makes reinforcements ambiguous and problematic. M#3 does not involve a visible Mamie, nor are there any participants making direct perceivable eye-contact (the crowd is at public

distance; footballers are internally focused on the game) which should be an ideal offer-offer complement with the lexicogrammar which is overtly exposition-based ('like many families here in Sierra Leone'). The overarching aim here appears to be a 'drawing back' momentarily to get a birds-eye 'offer' of gender issues in Sierra Leone. This is complicated however by two factors: first is the footballers' proximity to and kinetic vectors around the spectator's position (their use-space engulfs the spectator). Secondly, the crowd above looking down at the footballers establish their gaze-related use space to also engulf the spectator's position. The necessity for a more *detached* 'invisible onlooker' to objectively process the scene is hampered by the spectator's kinetic, spatial and gaze related implication.

Whereas the previous example points to a kind of CVR aberration, it is more common in the texts for there to be cross-modal incongruences between offer and demand. Also, owing to the complexity of the scenes it is not uncommon for shots to be both congruent and incongruent across the resources used. As I have pointed out, where the image presents a demand, it is invariably incongruent with the predominantly declarative verbal content. C#21 presents an obvious demand in the form of Ilungama Ayanda's gaze and forceful gestures toward the spectator. Visually he demands the spectator's attention and imagined response but verbally he is chronicling the suffering of his ancestors in declarative unmodalized MOOD structures. This assumes however that *he is the image*, and not just a component of it. At shot level, the construct becomes more complex: when the other pygmy men are viewed, they present more of an offer, as we can survey them neutrally without our gaze being implicated by their reciprocation.

C#25 involving Osambia-Kpwata Fyfy presents a more complex dynamic. In the shot, she is close to and 'alone' with the spectator and speaking in the diegesis (English, untranslated). Ostensibly a direct address, she is somewhat static, not oriented to the spectator and her gaze is obscured by her sunshades. Spatially she differs from the other highly demand-based direct addresses and instead presents a quasi-offer in her detached visual presentation (she 'offers' herself). It is tempting here to resolve the reinforcements by equating this as a cool and detached visual demeanour as congruent with her verbalised status-signalling ('chief of Zambia', Mobutu's daughter). In the same vein it would be possible to attribute Ilungama Ayanda's visual performance to the emotional content of his speech. To do so would however be to ascribe congruence *across functions* where the interpersonal and ideational functions are presenting an interpretation greater than the functional parts. This is not a problem in the broader analysis, where meanings are always cross-functionally overlapping. As a strictly interpersonal cross-modal analysis however it must be avoided, where analysis would 'build up' cross-modal inferences at a higher level rather than 'linking across' as is the intention of this portion of the

analysis.

I will revisit such cross-functional congruences instead as manifestations of *tenor* at the register level (chapter 8, section 8.3), where such intersemiotic congruence denotes situational characteristics such as familiarity, status, and social distance. There are other resources used in the texts that engender interpersonal cross-modal effects that will also be discussed as wider aspects of register (e.g., section 8.1.2 Medium), the most notable being the use of deictic reference ('*that* cross now filled with water' (C#28)). Deixis, as used in *Congo VR* references things in the image as if both speaker and listener (spectator) have direct access to those things. Deixis is typically categorised as a function of cohesion and will be discussed in that context, as well as its contribution to the texts' tenor, where it is described for its effect of engendering familiarity and social distance between speaker and spectator.

A substantial reappraisal of social distance, involvement, and power is required in light of CVR's matrix-like distribution of associations. As with Kendon's theories, semantics are located in a larger multi-directional syntagm than the hitherto bidirectional syntagm of framed filming. The example of the pygmy tribesmen highlights a fundamental concern for CVR as nonfiction representation: where multiple participants co-habit the shot, the spectator will often take a position with one, and away from another, and crucially where multiple heights are involved, the spectator *must* adopt an association with one or the other. The consequences here are two-fold regarding identification: first the spectator is blocked from identifying with the represented social actors and secondly, as in the case of the example given here, identification becomes an in- group/out-group dichotomy where the pygmies become othered by their stature relative to the BBC crew.

6.5 Concluding Remarks

This chapter foregrounded the nonfiction CVR spectator, for the ways they are constructed in CVRs immersive imagery. The spectator was found to be constructed at times in a tangible and stark manner and at others their construction was more nuanced. Both texts embodied the spectator as a feasible co-present human entity and augmented their perspective with socio-spatial meanings. Point-of-view is an important aspect of CVR, with theoretical and practical considerations. Through the texts, I have drawn out the complexity of CVR's POVs and use of perspective.

Multiple simultaneous subjectivities presented in both texts, where a stable objective view as found to be rare, if possible at all. The bystander construct was conceptualised as part of a wider flux of perspectives, shown to be relative both to other entities in the shot, and the field-of-view currently chosen. Like the transitive meaning potentials found in the ideational function, there are *perspectival potentials* available in the interpersonal. The continuum of views found included attempts at 'birds-eye' views, which I described for their problematic nature. I termed these as 'technical perspectives' in the texts, as they fall somewhere in between the neutrality found in conventional filming and the personified POVs in CVR, while fulfilling neither. At the core of CVR's perspectival complexity is that the spectator's reflexive 'real-time' view pervades all, including those views that are not intended to reflect an embodied position.

Contact was addressed as a problematised resource in CVR. As a by-product of perspective, contact is an unstable phenomenon where the 360° image naturally skews towards the spectator's socio-spatial involvement in the scene, negating the possibility for image-acts of offer. This is critical to understanding CVR from a textual standpoint as it requires a reimagining of the kinds of subject matter that are apt. Contact adds semantically in this regard to wider debates on the ethics of representation and modes of engagement with social subjects.

CVR's effects of simultaneity and multiplicity impact virtually all interpersonal subsystems. Where space-related social constructs were hitherto considered as bi-directional properties between viewer and subject (through the camera-proxy), they become subsumed in CVR into a matrix of relative associations. As the camera indexes distance and heights relative to filmed entities, so the rotating spectator does so in their perceived *social place* therein. This is fundamental to the production of nonfiction texts, where choices made regarding the placement and height of the camera carry semiotic risks that must be negotiated alongside the practicalities of showing necessary action (while not disorienting the spectator). I described the potential ramifications for representations of social actors and groups, where the spectator will

necessarily align and identify with certain participants, potentially forming an in-group, albeit in a subtle manner.

From a technical SFMDA perspective, I have incorporated Martinec's distance-involvement schema (Martinec, 2001) which he had developed in a non-textual context. More substantially, the introduction of Adam Kendon's use-space and formations added a level of descriptive clarity to the social complexity of the CVR image. Use-space was shown as a way of breaking the 360° image into socially (as well as transitively) active zones that incorporate other systems (proxemics, involvement, gaze, address for example) into larger analytic units. Kendon's Formations provided ways of making sense of the social arrangements in which the spectator will necessarily find themselves, when immersed in CVR. Considering the scene as a series of potential formations, adds to the issues of camera placement. The camera was at times found to be placed in-formation, where the spectator is most personalised, but at others the camera was placed seemingly for more practical reasons: close to the action or in between salient entities, where the positioning lacked a socio-spatial naturalness. Formations, as such, are a useful high-level guide for the placement of CVR cameras. From an experiential point of view, to have access to things like formations in shots adds to the immersive, social realism of the form.

While Kendon's transactional segments are semiotic and proved useful in conceptualising the interpersonal construction of the spectator in space, specific formations such as F-formations are more difficult to ascribe semantic functionality. They provide additional information to the multidirectional associative nature of CVR with resonances of contact, power, and perspective. The issue lies in their typological nature, where they have specific spatial qualities, and the spectator is either in a formation or not. Once analysis goes into more granularity it reverts to resources such as social distance and involvement. Also, an argument can be made for Kendon's formations and use-space being more textual in nature. Transactional segments, for example organise space in much the same way as framing does albeit based on purely social phenomena.

7. Information Organisation

The previous two chapters addressed CVR as a semiotic technology for the ways it configures fields of experience and social relations. This chapter uses the more inward focused textual metafunction, to see how CVR configures discourse as *information* in the 360° image and in its modal ensemble. The producers of CVR text are addressed for the challenges in controlling information: CVR presents unique problems when organising entities within and across 360° shots, where framing would otherwise allow for things to be selectively excluded. The findings are presented as the results of a visual analysis of the texts' shots, as well multimodal macro-analyses of phases from both films.

7.1 Framing

CVR's lack of a frame has been implicit in much of the discussions thus far, with the resulting effects of multiplicity and simultaneity foregrounded. This section addresses the 360° image directly, for the way it organizes visual information. The texts were analysed for the ways that the image connects and disconnects visible entities, in ways described by Kress and van Leeuwen (2006) and consequently for ways that the 360° image reformulates conventional framing. CVR's novel modality adds the perception of a ubiquitous ground-plane and a sense of volume in the mise-en-scene surrounding the spectator. This added dimension frames by connecting and disconnecting entities in three-dimensions. Connected to this, framing necessarily involves the spectator, who is at all times framed in or out of larger information arrangements. The discussion will mostly account for the organization of visual information at the rank of *shot*, for how the mise-en-scene is constructed informationally.

As brief preface to the discussion, I will compare the CVR camera with conventional cameras for their capacity to use abstraction as part of framing. Unlike conventional cameras, CVR is unable to exclude via the frame, and unable to optically abstract the forms that it does include. Abstraction is achieved in conventional photography through manipulation of depth of field, such as in Cartier-Bresson's portrait of Picasso (1953), seen in Figure 19, below where the subject is framed within a partial, defocused archway form. Here Picasso is disconnected, isolated visually and figuratively, adding poignancy to his gaze.
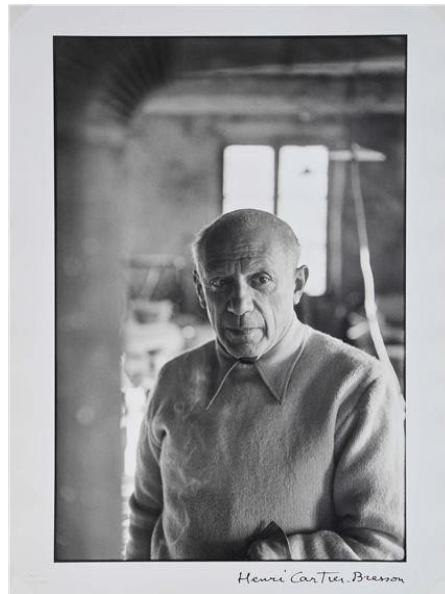
Figure 19: Pablo Picasso (Cartier-Bresson, 1953)

Importantly here, it is the inability for the viewer to resolve the archway form beyond its framing function; its details are obscured, and 'rotating' will not reveal anything beyond what is shown. CVR cannot generate this level of abstraction for two reasons. First, the optical mechanics of the CVR lens means that it cannot create shallow focus and thus loses the capacity to abstract through shallow depth of field. Secondly and more fundamentally, the rotational aspect of CVR means that any potentially abstract shape will be immediately resolved by the rotating spectator into a recognizable object. In chapter 5, section 5.1.3, I addressed this as the ubiquity of visual setting (circumstance) in CVR images. Transitivity is thus 'at risk' in the textual function in a particular way in CVR where shapes become windows and edges become walls; a phenomenon that points to a reconfiguring of visual discourse.

Given that the 360° camera makes available all visual information in all directions, the mise-en-scene was scrutinized for the ways it manifests frames, as lines, shapes, surfaces, and negative space. The availability of the ground-plane results in information proliferating in all directions simultaneously, with varying degrees of 'strength'; left-right, up-down, and towards-away from the spectator. Certain types of frames emerged in analysis, allowing for a loose categorization of frame types.

- *General frames*, where conventional framing devices were implemented on the vertical plane (in other words, away from the ground-plane)
- *Stage*, and *container frames*, where frames were on or emanating from the ground-plane.
- *Windows*, frames as junctures between the visible 360° mise-en-scene and an

imminent outer world beyond the spectator's purview.

General frames are less novel than the others but notable nonetheless for how the 360° space can be divided up informationally using conventional means. Rectangular shapes provide clear frames throughout; schoolrooms for example include a blackboard that delineates the teacher and their activities (and status). The doorway in M#4 (Figure 20, below, right) provides a frame, not unlike Bresson's, intended partly to show her isolation. Another disconnecting frame is shown below (left), where the clay incline in M#3 forms a horizontal dividing line between the footballers and spectator below and the pupils watching above. While this involves a ground-plane, it can be considered two-dimensional as, without a line-of-sight to the platform above, the line splits the image vertically, creating a two-dimensional barrier between watchers above and watched, below. This kind of arrangement will become important in other informational systems discussed below, that carry their two-dimensional functions into CVR but do not function 'natively' as 360° effects.

Vertical disconnecting frame (M#3, 1:01)   Overlapping disconnecting frames (M#4, 1:43)
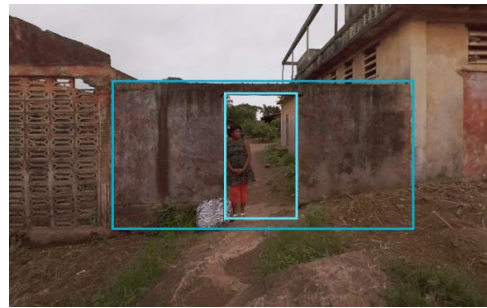
Figure 20: Examples of general frames in Mamie's Dream

When the ground-plane is considered, the spectator, embodied by the situated camera is incorporated and framed by default, and where *containers* and *stages* are also found to connect and disconnect them with entities. Containers are at times literal, as in the boats in M#1 and C#4. In both shots, there is an obvious inside-outside separation of information and in both cases the spectator is incorporated within the frame. In the case of C#4, the boat frames the spectator as part of 'the team and I' establishing a compositional unit and disconnecting it from the large expanse of river around it (already informationally loaded as 'beautiful but troubled'). Here the shape of the boat goes further to internally organise the boat's inhabitants as the lines of the pointed bow converge sharply behind and the frame also 'point' to Alastair Leithead., reinforcing his status, established across the visual and verbal modes. In Mamie's case, the boat is doing textual work, by connecting the spectator to a yet unspecified character and disconnecting both parties from the unknown surroundings. Where the boats provide container frames that are used to include, e*xclusory* framing was found to distance and isolate the spectator. M#2 (on the bridge)

144

presents a complex series of frames, the principal one being the bridge which is both *stage* as a flat platform, and *container* when the railing is included. Here, Mamie and the spectator are close but disconnected by the railing. The bridge, as stage, is clear in its function; to enhance Mamie's outcast status through negative space. The barrier, as container however is ambiguous, as it disconnects the spectator from Mamie, reversing the effect of the previous shot.

C#13 provides a stage frame in the form of a raised footpath that forms a clearly bounded arena for Junior to enact his verbal address to the camera-spectator (Figure 21, below). The frame here also serves to delineate all those ratified as part of the production, and those excluded as generic (a disparity reinforced by the height of the footpath). Beyond the ethics of the frame, it is also problematic in a more basic informational way regarding the woman who shares the stage with Junior and the spectator. Here framing points to her ambiguous status as *framed-with* junior, who is unambiguously presented as important. By virtue of inclusion in the framing, she is presumed to have status also that is not specified thereafter, a point discussed further at the end of this chapter.

Figure 21: Example of a stage-frame (C#13, 3:05)

Finally, *windows* are included as an additional way of graphically connecting and disconnecting information but do not operate in the same way as the previous examples. The overarching concept with this kind of frame is that they provide a bounded plane through which participants can engage with information that is imminent but not visible in the shot. Doorways and windows naturally lend themselves to this effect. The doorway in M#4 provides an inside-outside juncture between the visible *public* outside and the obscured *private* inside. Here the disconnecting is both the isolating of Mamie within the door frame and the partially visible separation of internal-external information (potentially supporting readings of her ostracization). Actual windows are found to serve a similar purpose such as in C#8 where Alastair Leithead is gazing out of the train

window at things the spectator cannot see (Figure 22, below). These are notable from a textual point of view as they are perhaps the closest equivalent of a subject looking 'off screen' where otherwise, gazes tend form vectors within the shot, directed at other participants. In a sense, by providing windows such as these, the gaze vector is neutralized by not implying a target, and as such do not place additional transitive meanings at risk of signification.

Figure 22: Alastair Leithead looking 'off-screen' through a window-frame (C#8, 1:50)

7.2   Salience and Reading Paths

Previous chapters discussed 'entities' in the form of transitive participants; figures and episodes engaged in processes. These were assumed to be sufficiently prominent to have a status in their field of activity. Similarly in the logical function, the components being linked (i.e., *conjunctively relatable units* - 'CRUs') were assumed, prior to their conjunctions with other CRUs. Salience is the compositional system that tracks the ways otherwise circumstantial elements in the scene become foregrounded for their specific semantics. A linguistic SFL equivalent for salience is *theme*, which marks out a lexical item for its prominence in the clause. In this light, the texts were analysed for the ways that visual entities became thematic, as *visual topic* through figure-ground contrasts of scale, colour, shape, and movement. Once established, salient entities were then considered as dynamic points of interest (POIs), as nodes threaded along a reading path (RP) that feasibly represents the spectator's attentional vector through the film. This aspect of analysis forms a partial basis for further consideration of the texts' *cohesive properties* (cohesive ties and conjunctive relations) that are discussed separately in section 7.4, *Textual Cohesion In CVR*. As such, I will not discuss how the spectator makes sense of the reading path, rather describe the potential resources involved in organizing the path itself.

Scale was a significant contributor to salience: Mamie, Alastair Leithead, and the other speakers in *Congo VR* were salient as the optical consequence of their proximity to the camera. This is relatively uncomplicated as a natural resource used to denote salience. *Colour* manifested salience in places such as with Alastair Leithead who is salient in C#20 as the only white-skinned

146

participant in the shot. *Movement*, is considered as a strong determinant of salience (Itti, 2005)[25], but was not found as a sole factor of salience except in a reflexive sense in M#13 where Mamie is walking through the large crowd of people, tracked by the camera's movement. In this instance, Mamie is salient by her and the camera moving in sync, but with the result of her appearing optically static, while the others in the scene recede away. In other shots, movement tended to coincide with scale such as where Mamie is grooming her daughter (M#7, 9) and in the forests where the pygmies are digging, hunting, and gesticulating forcefully towards the camera (C#21).

Where scale, colour and movement provide the potential for strong salient features, other less pronounced kinds of figure-ground relationships were found in the scenes manifesting as *contrasts of shape* and form. Geometric, manmade forms contrasted with organic backdrops, such as in the tracks and pits etched out of the DRC landscape in C#3 and the 'cross now filled with water' in Mobutu's overgrown ruins C#28. Finally, salience and framing merge in places through negative space, where entities are salient as contrastive with their environments. This is found in many shots where a direct address is used but is particularly strong in C#25 where Osambia- Kpwata Fyfy is salient through variance against the negative space on the roof deck of the ruined palace.

The 360° image presents difficulties for producers in maintaining and managing salience in the shots. Where scale is used most frequently to demarcate those who are topical from their circumstantial surroundings, it is occasionally disrupted by smaller, more salient moving entities. Mamie's static demeanour in M#2 and its conceptual overtones are briefly interrupted by the motorcyclist driving across the bridge, affecting the overall dynamic of the shot (0:39). Movement is also a significant factor in complicating the conceptual readings of *Congo VR*'s aerial shots where the drone is seen clearly above the spectator. Other complications are choice related. Salience as size can be used erroneously and mislead the spectator's attention such as with the woman sharing the stage with Junior in C#13, who has salience but no significance in the overall reading path.

CVR's simultaneous flux influences the perception of salience. Tseng's systems of *immediate* and *gradual salience* (Tseng, 2013) proved useful for analysing CVR's spatiotemporal mix. CVR's differences to framed film are highlighted when considered against her systems, where for example 'immediate salience' (salience established at the shot's commencement) becomes less of a fixed property. 'Immediate' implies that the spectator will have the entity in their field-of-

---

[25] Serving a valuable protective and evolutionary function (our most peripheral vision is attuned to movement (Mckee and Nakayama, 1984).

view but owing to the spectator's continuing rotation, only the very first frame in a CVR film is spectator's field-of-view assured (the viewing software is designed to ensure that the film will begin 'on front' of the spectator regardless of where they are looking). In any frame thereafter while it is often probable, it is not guaranteed, leading to other kinds of salience. Connected with this, the 'gradual' nature of Tseng's salience must also be reformulated to account both for objects becoming salient through their own spatiotemporal properties and for the spectator's rotation *revealing* salient entities. I will address the nuances of dynamic saliences here as part of CVR's phenomenon of simultaneity. For reasons outlined above, I will avoid the term 'immediate' and instead organise my discussion around three types more apposite for CVR: *simple*, *gradual*, and *split* salience (illustrated in Figure 23, below). To account for the dynamic nature of CVR's rotation, I will also discuss it for the way salience interacts with field-of-view.
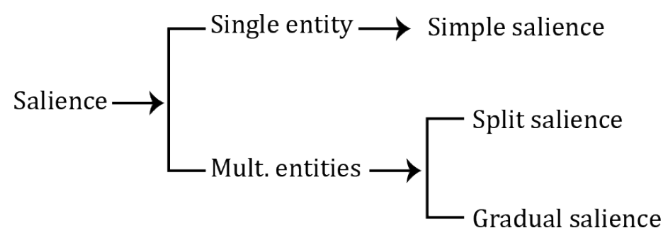
Figure 23: System network illustrating choices of salience

Simple salience is so-called as it denotes situations where there is a singular salient entity present throughout the duration of the shot and where no other salient entities emerge. *Mamie's Dream* presents several shots where she is salient in this way. Her proximity and scale in the boat (M#1) is straightforward; her compositional units in the bedroom (M#7) and on the school steps (M#17) are similarly so. Congo presents scenarios where a subject is multimodally salient against a relatively homogenous backdrop. The Pygmy man digging (C#19), provides simple (albeit weaker) salience in this regard. As there are no competing elements in these shots, rotation can be typified as *surveying*, where the preferred reading will necessarily lead back to the salient entities, having found no others. This can be contrasted with the next kind of *gradual* salience found in the texts, where rotation is at times motivated by the characteristics of a gradually emergent salient object.

Gradual salience concerns entities becoming more, or less salient over time as they emerge from, or enter, more circumstantial statuses. This takes many forms, mostly determined by movement which can be participant-based or camera-based. *Congo VR* is replete with shots where people move towards and away from the static camera becoming more and less salient. C#10 involves a stream of generic participants culminating in a moment of stark multimodal salience when

woman moves across the camera at intimate proximity (salience of scale) and shouts towards an unseen person. In C#18, pygmy tribesmen move from the foliage into, and out of view, gradually gaining, losing, and 'swapping' salience. Where camera movements are involved, relocating the spectator towards and away from entities, perception of their scale makes them gradually salient. In M#4 Mamie is the target of a slow tracking shot through the playing girls and again in M#19 where a much longer vector brings her into salience (6:57) momentarily before ultimately moving far enough away in its movement that Mamie is reduced in size merging into the surrounding riverbanks (7:03 onward).

Salience has two different functions in the latter example (M#19): the first serves to gradually realise Mamie's visual weight of importance, the second serves to diminish Mamie in so far as her story is resolved and she no longer requires foregrounding. This kind of salience enables the transitive flux described in the chapter 5, where circumstance and process merge and exchange entities. The significant difference between this kind of salience and the *simple* kind described above, is that field-of-view is now semiotically constrained by the kinetic-temporal space of the video representations. As kinetically motivated (at least to some degree), the spectator's rotating movement and the scene's filmed movement are two overlapping kinetic systems that coalesce to determine the attentional vector through the shot.

The third kind, s*plit* salience involves entities with comparative levels of salience co-existing simultaneously at shot level but not within a singular field-of-view. As such, the spectator's rotation is motivated similarly to gradual salience but in this case the rotation is more subjective where the spectator encounters and resolves diametrically positioned entities. There are logical aspects to this configuration as split salient entities will necessarily be compared for meanings to be made. Salience is split in *Mamie's Dream* between the teachers and pupils (M#6, 13), who occupy opposing sides of the image. In Congo, the Tribal ritual in C#17 places the spectator between the tribesmen and the BBC crew, and as such splits the composition between two salient episodes. Split salience can morph or 'dissolve' in shots where there is movement. M#3 combines salience types as it begins with a split salient composition where the footballers and the crowd oppose each other in the space; the footballers' activity then crosses the spectator's position midway through the shot and leads the spectator to the crowd.

Salient entities were considered for the way they form points of interest (POIs) along the texts' *reading path* (RP). CVRs omnidirectional image and long shot durations,[26] present unique challenges in characterising RPs. As an edited sequence of long duration 360° shots, the reading

---

[26] *Mamie's Dreams*' shots range from 16-35 seconds with an average duration of 23 seconds, *Congo VR* ranges from 10-31 seconds with an average of 19.

path will manifest both around the image as an *intra-shot* RP, and across edit points presenting *inter-shot* RPs. Intra-shot RPs are constituted by any vectoral means potentially motivating the spectator to redirect their field-of-view. This manifested in the texts as moving salient entities, moving camera, transactional vectors, social gaze, and geometric elements such as leading lines and frames. Inter-shot RPs were formulated using Brillhart's seminal work on VR editing (2018b) and her *shot matching* was found to be a significant device used to maintain the reading path across edit points.

Having made a preliminary pass through the texts, I observed that salient entities tended to match across the edit with other salient entities. Using Brillhart's 'probability' terminology (Brillhart, 2018b) I termed these matches as 'end probability frames' (EPFs) in the last frame of outgoing shots, and 'start probability frames' (SPFs) as the first frames of incoming shots. Shot matching in this way assumes two things: that the first frame of the film's first shot contained a 'guaranteed' SPF that could be considered as the first point on the path, and that there would be no semiotic basis (or time for that matter) for the field-of-view to change mid transition between shots.

A synoptic, intra/inter-shot analysis was performed to track where RP resources joined-up into a continuous flow of action, what Kress and van Leeuwen refer to as a 'plausible reading path' (2006, p.25). The objective of the analysis was to see where the resources cooperated, maintaining direction through the texts, and where the path broke down. Figure 24, on pg. 151 shows an extract from the full reading path analysis of *Mamie's Dream* (full analysis is found in Appendix J. Reading Path Analysis. Mamie's Dream) with visualized examples of inter- and intra-shot RPs, as well as the vectoral resources used to direct the spectator along paths.

The clearest matches were found between *Mamie's Dream*'s shots 1 and 2, as seen in Figure 24, below, where she is clearly salient in both and perfectly aligned in the same field-of-view across shots. A synoptic process is shown thereafter establishing POIs at the start and end of shots and connecting them across the edit. With a series of likely POIs established, the intra-shot RPs could be evaluated based on the attentional vectors enumerated above. Where a shot suggested a divergence of a path (such as in split salient compositions), both were considered to lead to their respective EPF, suggesting branching paths with multiple SPFs in immediately following shots, from which further RPs potentially emanated. An example of a *branching RP* is shown in Figure 24 between M#2> 3. Shot two takes Mamie as its SPF (matched from the previous EPF of her in the boat) and from this point, her gaze, directed out onto the river behind the spectator, prompts a rotation of view. The river's converging on the horizon becomes one of two EPFs (Mamie herself being the other).

Shot 4

16. EPF2: If movement is primary RP motivation

15. EPF1: girls playing in slow motion - If target of M gaze is RP

14. RP: M gaze (with head move) directs focus to girls playing

13. RP: M becomes larger and salient throughout movement – have access to her gaze

12. RP: camera movement leads focus

11. SPF: Follows kinetic vector of ball - in previous shot

Shot 3

10. EPF Follows strong kinetic vectors picked up from either SPF -RPs

9. RP: as continued from SPF2 - OR as directed from onlookers' gazes in SPF1

8. SPF2: lines up with river in previous EPF

7. SPF1: Matches M in previous

Shot 2

6. EPF2: When RP follows M's gaze vector

5. EPF1: When focus remains on M

4. SPF: Matched with previous view

Shot 1

3. EPF lines up with Mamie in next shot

2. SPF: perspective given at onset

Titles

1. SPF: perspective given at the onset

Figure 24: Extract from reading path analysis of Mamie's Dream, shots 0-4

151

An inherent danger in the approach was the potential for subjectively engineering intra-shot RPs to 'fit' the more readily found inter-shot matches, essentially shoe-horning the RP to suit the process. To counter this, analysis allowed for an unlimited number of diverging (branching) paths, wherever they emerged, and to accept the possibility of 'broken paths'. Another methodological concern in this approach was that the RPs would branch exponentially into unmanageable branching paths, but this turned out not to be the case. *Mamie's Dream* branched into a maximum of two distinct EPFs in different fields-of-view.[27]

RPs were found at times to be problematic, 'weak' and at times 'broken' where no clear EPF-SPF match to be made across the edit. An example of this is the un-established or resolved parked motorbike in M#5 (see full analysis in Appendix J. Reading Path Analysis. Mamie's Dream p.5 for inter-shot context). Through salience (size and shape) and composition (a triangular arrangement connecting the vehicle by the spectator and Mamie) the motorcycle becomes a textual 'red herring', seemingly a participant in an imminent process. Its salience nonetheless affords it possible POI and as such EPF status setting up a curious series of matches across the edit: the motorbike matches with the pupils as SPF in the next shot, and Mamie matches with an empty schoolroom doorway. As such, the RP becomes weakened by either an unresolved expectation of *Mamie's* intended SPF (the teacher emerges through the doorway only after five seconds), or the potential for the spectator to peruse the many pupils spanning the room, and potentially miss the teacher and his subsequent transactional vectors.+

Figure 25: Unresolved salient object (motorbike) leading to ambiguous reading path

Where these examples point to a problematic EPF, other appear to terminate altogether. In the

---

[27] The illustrated analysis in Figure 24 shows two EPFs ending shot 4. The full analysis includes a third EPF, where two are marginally different and in the same field-of-view.

transition between M#15 and 16 the path appears to break down at the SPF stage of the second shot. In M#15, Mamie and the girl are clearly salient by size, shape, colour, and as matched camera movement. They form a likely EPF which is followed by the SPF of the girl, without Mamie. The issue here is of the girl's weak/non salience in M#16 where she is not sufficiently 'variant against invariant' among the crowd of similarly dressed girls. This mismatch arises as she is already specified (singled out) in the previous shot (*walking with Mamie*) and now appears circumstantial. In a general continuity sense, her demeanor is also partly responsible for the breakdown of the reading path; where she was relatively downbeat in the previous shot but is now seen eating and dancing. These issues reflect the materiality of the continuous flow of visible information across the image and shots. The semantic repercussions are addressed as *cohesion*, later in this chapter.

7.3 Information Value and Textual Rhythm

The texts are discussed here for the myriad ways that information takes on characteristics of *value*, dependent on how they are positioned relative to each other in the image. Kress and van Leeuwen's spatial information regions, listed below, were applied in the texts to ascertain their relative usefulness, and for how the immersive modality alters their use. Martin and Rose's linguistic information structuring, as *waves of information* (2008) is included as a reference frame for the temporal aspect of information structuring in the texts. Information value, as per visual grammar was tested in the texts along the following spatial dichotomies.

- *Left-right*: given and new
- *Above-below*: ideal and real
- *Centre-margin*: weighted importance and peripheral lesser-values (more quant)
- *Front-back*: three dimensional diametric values (more quant)

As with salience, CVR presents fundamental challenges in applying fixed values to entities that will shift spatially, dependent on the rotating field-of-view. Visual information value was originally conceived on the page and screen which are fixed and not dependent on the viewer's point of view. In CVR, 'left' and 'right' are not absolute values when the spectator may shift their view, rendering an object either as *left*-to object X, or *right*-to object Y. Where entities are diametrically opposed in the shot, they can be either left-right or right-left depending on the field-of-view chosen. Paired with this, it also runs counter to human vison to expect the spectator to keep a participant on the one 'side' of their field-of-view, where the inclination will be to rotate and to centre them in the view. In other words, we do not 'see images' in CVR as gestalts as in

visual grammar's original information value systems. Taking these overarching constraints in mind, information values were attributed as following.

Left-right, *given-new* dichotomies present in both texts. *Congo VR* uses given-new information to set up screen-direction and continuity between locations. This is evident in the sequence of shots involving the train journey (C#7, C#8, and C#9). In C#7, the busy train station is composed in two halves where the station building is on the left and the train is on the right.[28] A left-right dichotomy establishes the station as *given*, as a stable entity and the train as *new* or that which *emerges* from the station (of course, train stations are also places of arrival, but this would simply be a different kind of new). This left-right configuration holds for the following shots where the more salient window is to the right, through which Alastair Leithead extends his gaze to an external and *new* - inland DRC.

In M#7, Mamie is left and given, and her daughter is new. Compositionally, left-right is visually readable as they are alone together, and close enough to make the left-right connection in a single field-of-view. Here, her daughter has a temporal *new* status as it is her first appearance in the film and a conceptually new status as 'the message' of the shot (Kress and van Leeuwen, 2006, p.180). In this sense the girl also adds a *new* dimension to Mamie, conferring a maternal quality on her. The direction flips in M#9 however where Mamie holds her daughter between her thighs appearing as daughter (left) > Mamie (right). As part of the wider shot, they are both left and right relative to other entities in the scene. In this regard, Mamie can perhaps be *new* to the given readers where they are considered the norm/status quo. This dynamic resembles Martin and Rose's informational waves in that the earlier shot culminated in a wave of information cresting with the Mamie's *new* value (left > right as Mamie > child/motherhood), a value that becomes incorporated into a subsequent larger wave of information. Essentially, once the first shot does its semantic work, it provides the resources for the culmination of larger meanings (and so on, upwards). Here, Mamie (and child) must be considered a *new* element to the *given* readers to her left. The nested wave of information is visualised below.

---

[28] This assumes a reading path EPF 'bringing' the spectator to this composition. There is a supportive EPF in shot 6 in the form of the map of the DRC. Regardless, the relations are discussed here as abstract concepts and textual potentials and not intended to necessarily fit into an overarching reading of the film.

Big wave (Shot#9) --▶

Small wave (Shot#7) ---------------------▶

*G:[G: Mamie N: daughter = maternal Mamie] **N:** readers = Mamie's maternal social effect*

Above-below information values were found in places to faithfully replicate visual grammar's ideal-real concepts, but only where there was a clear compositional split between the upper and lower parts of the mise-en-scene. M#3 is a prototypical ideal-real construction as it is comprised of a 'clean' vertical split made by the straight horizontal line at the top of the clay incline, separating the onlookers above from the footballers below. The result can be construed roughly as: *below* the footballing boys are highly kinetic and chaotic in their vectors, exhibiting unfettered, 'real' boyhood while the onlookers above are static and 'idealized' by their placement on a platform, 'raised up' literally and figuratively above the fray. A similar effect is achieved in the later shot M#8 where Mamie is positioned on a raised platform looking down onto the market below. The market traders are both spatially below and homogenised temporally and as such 'real', as *quotidian* with Mamie on the raised porch idealized. This shot differs by degrees however, where she is not as high as those on the clay incline relative to those below and inhabits a different field-of-view. As such there is no clear vertical contrast between her and the traders and lacking the *visual polarisation* which is considered foundational to the construct (van Leeuwen 1998 in Stenglin, 2009).

The centre-margin information system was found to be a feature of certain scenes feasibly contributing to the overall structuring of information as relative values in a radiating formation around a nucleus of importance. M#13 presents one such configuration where Mamie is portrayed as the nucleus of the scene and all others are in the margins around her. Here the configuration is marked by the movement of the camera, which moves in sync with Mamie and by doing so keeps her centred and relegates all around her to the margins. A static example can be found in C#21 where the pygmy elder is centred, framed in his stage, with the younger tribesmen in a loose ringed formation around him. Here the compositional aspects of the shot reinforce the marginality of the other tribesmen. M#9, already discussed for left-right, given-new structures can also be considered as centre-margin, albeit in a very loose fashion. Given that she is nestled between the readers on her left and the others on her right, she is also *centred* within them and a fulcrum of sorts to the others seated in the more distant margins, off the porch area. Here her status is highlighted by her centrality.

The last information system, *front-back* was used to analyse information values in the three-

dimensional space. The system is predicated on the status of information as it presents in volumetric contexts but is problematic in the conceptual distance required from the original concepts to their potential implementation in CVR. The system as intended by van Leeuwen (2005) describes the value of labelled bottles, buildings, and other designed volumetric entities for how their 'face' and 'rear' have informational value equating the objects to the human body (as literal face, and where we expel waste (ibid. p.211)). To apply the system to human participants in CVR becomes tautological, applying it to the literal entities that the system symbolises. Secondly, *volume*, such as bottles and buildings. are available to the viewer *in their totality* when rotated or circumnavigated, whereas objects in CVR are available only for the parts that face the camera. It is possible to attribute the values to participants such as the dancing man at the church event in M#14 who remains oriented away from Mamie and the spectator's position, throughout the shot. The effect here is also interpersonal, of reduced involvement with Mamie (he is 'showing her his back'). In a very different way, Mamie's private talk with her pupil (M#17) takes place at the back, or side of the school which by the system's reasoning also infers value but how that value can be characterised and operationalised in analysis is difficult to gauge.

The information systems devised in visual grammar have merit in the analysis of CVR but require re-evaluation as to where they are most useful. Certain shots it the texts lent themselves to conventional value readings such as the above-below information provided in M#3. It seems however that such uses are predicated on 'flattened' compositions such as this one where there is a clear polarisation of what is above and below. When the subsequent example (M#7) was compared with it, the spectator's embodied position becomes a factor in the verticality. the overall textual usefulness of the information zones seems somewhat diluted in their uses in a heavily perspectival form such as CVR. When the embodied spectator is factored in, 'above' becomes enmeshed in the interpersonal aspects of power and association. As such while the systems are useful, they may not be useful as CVR semantics but rather usable where CVR resembles framed compositions.

7.4 Textual Cohesion In CVR

Halliday and Hassan describe coherence as the way texts 'hang together' (1985, p.48). This section addresses question one directly for the way the CVR texts 'hang together' as meaningful *coherent* multimodal experiences. This extends the intersemiotic meanings discussed in the ideational chapter, to include intra-visual and cross-modal interconnections extended across macro units of text. The section is broken into two subsections discussing the texts for the two

kinds of cohesion found in texts, *logical conjunctions,* and *cohesive ties.* In each subsection, I will first compare the texts at a high-level, and then present the findings of detailed analyses of a phase from each text.

7.4.1   Conjunction

The analyses of conjunctive relations between CRUs are multidimensional in nature. For the necessity of describing it here linearly, I will use the following sequence to describe relationships.

1. Between visual CRUs within and across shots
2. Between verbal CRUs (clauses) within and across shots
3. Between visual CRUs and verbal CRUs within and across shots
4. Between cross-modal logical meanings culminating from 1, 2 and 3

For reference purposes, I include a visual of the conjunctions in the form of a reticulum and a key to the CRUs used, (Figure 26, p. 161) an approach taken by van Leeuwen in his multimodal analyses of documentaries (see 1991, p.97).

At a high level, conjunction in the texts reflects the mixing of narrative and exposition, where the texts cohere logically in their construing of the experiential flow of narrative spacetime and their explaining of matters through comparison and causation. These reflect the films' generic mix which are elaborated in detail in the next chapter. Essentially, *Mamie's Dream* presents a narrative in the verbal track which is contained within a larger report. Mamie's back-story ends in M#11 and the text switches from historic biography to compositional report, where the Learning Assistance Programme (LAP) is elucidated and decomposed for its benefits (see Figure 29, p. 185). From a conjunctive point of view, this switch is signalled in M#9 through a *hyper-conjunction*, where the second half of *Mamie's Dream* expands the first through a causal conjunction of *purpose.* The first half of the film 'causes' the second and this conjunction occurs between M#8 where she is 'alone' in the market, and M#9 where Mamie verbalizes a moment of realization, 'I sat down and thought to myself, I am not going to give up on my dreams, I am stronger than this' - followed by a new phase and upward trajectory in her story. As such there is an implicit conjunction equal to 'for that purpose' or 'with this in view' (van Leeuwen, 1991, p.80) connecting the film's halves.

At lower levels, Mamie's verbal conjunctions take the form of additive extension and temporal enhancement, to incorporate a disparate range of participants and to establish the episodic historic timeline. Within shots, Mamie connects her spoken clause complexes as mini causal

threads that are resolved before moving on:

'I refused [Result >] I was scared I might die [< reason] I was twelve at that time [< reason]'

'My father said I was a prostitute, [< result] so he drove me and my belongings out onto the streets. [< result] I was alone.

Congo VR presents a series of implicit *elaborating* conjunctions, at its highest level, with social actors selected to *particularize* aspects of its socio-cultural thesis (that the DRC is beautiful but troubled). Where Mamie represents history through her own voice, *Congo VR* moves the camera-spectator through the DRC allowing social actors to represent their respective themes. As a multilocational film, Alastair Leithead uses contiguous spatial qualifiers to move the spectator from location to location. Examples include 'The first of the Congo River Rapids, *at sixty miles from the sea*' (#6) and 'Far from the capital' (#16), which are spatially contiguous with earlier phases (the visible sea, train station and traffic). 'Kinshasa' as a non-contiguous spatial qualifier is elaborating through ellipsis: '[*this place* we have arrived in *is called*] *Kinshasa'*. Other macro-conjunctions were more expository in nature as replacive ('on the contrary'), and comparisons ('alternatively') such as 'Despite the difficulties of living here, people do exciting and surprising things', which is a macro-conjunction that extends the 'difficulties' represented previously (from the rapids, through train-travel, to the busy traffic) and sets up Junior, Mama and the Sapeurs as triumphing over adversity in 'surprising' ways.

7.4.1.1   *Mamie's Dream*: Phase Analysis

Mamie's Dream was analysed across shots one to five, a portion of the film constituting a bounded phase where Mamie completes her back-story, adds cultural context, and the inciting incident(s) and conflict that propels the ensuing narrative. The CRUs are listed in the table below; their logical connections are illustrated graphically as reticula in Figure 26, p. 161.[29]

---

[29] For reasons of space, the CRUs and the reticula are separated in this document. They are available for side-by-side viewing in the Appendix I. Conjunctive Analyses.

Shot	CRU	Visual	CRU	Verbal (incl. explicit conjunctions)
#1	V1	Mamie in boat	1	My father said
	V2	Man rowing	2	educating a girl is a waste of money
			3	When I was a little girl, I told my father and mother,
			4	my dream was to become a teacher
#2	V3	Mamie in bridge	5	My father said, "
	V4	River	6	girls, they just get married and move away"
#3	V5	Footballers	7	My parents loved their boys
	V6	Crowd	8	(+) Like many families here in Sierra Leone, my parents wanted their girls to join the Bundu society and enter into womanhood
#4	V7	Mamie	9	(+) I heard
	V8	Girls playing	10	it's painful
	V9	Girls slowed down	11	(=) They blindfold you and cut your clitoris
			12	(=) It's called FGM

Table 13: CRUs for conjunctive phase analysis of Mamie's Dream

Explicit conjunctions in the phase are predominantly temporal, additive and projection-based, but when the verbal and visual CRUs are combined into a cross-modal logical complex, the phase shows its underlying expository function involving gender discrimination and cultural values.

Beginning with the spoken conjunctive units, CRUs 1-6 are a sub-phase comprising a series of verbal projections by Mamie and her father. As well as projecting clauses, CRU 3 *qualifies* the other CRUs temporally ('when I was a little girl…') and CRU 6 "Girls, they just get married and move away" *qualifies* CRU 2 "educating a girl is a waste of money", as *cause/reason*. This kind of rhetorical conjunction is a means of explanation, which reappears in the next phase where she 'refused' [because] she was scared [because] she was just twelve years old. Verbal conjunctions are also used to *elaborate*, by distilling (*particularizing*) 'it's painful' (CRU 9) into the specific action of '…cut[ting] your clitoris'. This is immediately followed by a *summative* distillation, where she sums things up with the label 'FGM' (CRU 11).

A more implicit logical construct appears between CRUs 7 and 8, allowing for a higher level of exposition where *girls* and *boys* are contrasted verbally as a means of characterizing parental inclinations towards gender, and the Bundu society itself. The clause 'parents wanted their girls' is an implicit negative comparison with 'parents loved their boys' ('on the other hand') where the common mental processes *loving* and *wanting,* provide adequate grounds for direct comparison

of 'boys' and 'girls'. This boys-girls dichotomy is reinforced in the visual conjunctions, and consequently cross-modally (discussed further below).

The visual conjunctions within the shots largely are of *kinetic contrasts*, between Mamie (V1) and the rowing man (V2); the footballing boys (V5) and the standing crowd (V6), and between Mamie seated in shot five (V7) and the flurry of playful activity that flanks her on both sides (Vs8,9). In terms of visual conjunctions between shots, the most consequential relation is found between the footballers (V5) in shot 3 and the girls playing in the next shot (V8). This conjunction approximates to an alternative or *replace expansion*, boys play football [and on the contrary] girls play hopscotch. This can also be interpreted as a negative comparison (by contrast...), but regardless of the exact conjunction, the relationship marks out boys and girls as different in some way which forms part of a larger logical complex when conflated with the verbal track. The conjunction of spatial contiguity is found between shots 1 and 2, as they are both on the river near/on the bridge.

When the visual and verbal CRUs are combined, the phase becomes logically cohesive through the emergence of higher-level inferences. The verbal CRUs 7-8, were described above as implicit negative comparison where *'loved their boys'* contrasts with *'wanted their girls'*. Cross-modally, the visual footballing boys (V5), *exemplify* 'their boys' (7). By extension, their contrastive counterparts watching from above take on the logical cross-modal association with 'their girls' (there are some boys watching too, but the overall effect is akin to a passivation/feminization). The inter-shot construct between 3-4 adds fundamental logical meaning where the visual *playing girls* in shot 4 (V8,9) are visually contrastive with the previously seen *footballing boys* (V5) while also cross-modal exemplifications of the verbalised '*their girls'* (8) from the previous shot. A criss-crossing chain of conjunctions emerges where the girls seen in shot four are the culmination of *boys/girls* equating to *loved/unloved*, and where *unloved girls* join the Bundu society.

Of note here is the way this chain takes on subtle conjunctive aspects where their play is also, by inference, *enhanced* by circumstantial features of *manner* and *role*: they play *happily* (manner) but *as girls* (role), their visualized play belies the quality/disparity of their prospects (Bundu being conflated with 'womanhood'). The effect is completed in shot four where Mamie is also represented as visually contrastive (static/kinetic) and her verbal conjunctions ('it's painful' > 'blindfold you and cut your clitoris' > 'FGM') form a high-level *negative comparison* with the carefree girls, whose play is distorted in the final frames through a slow-motion effect.

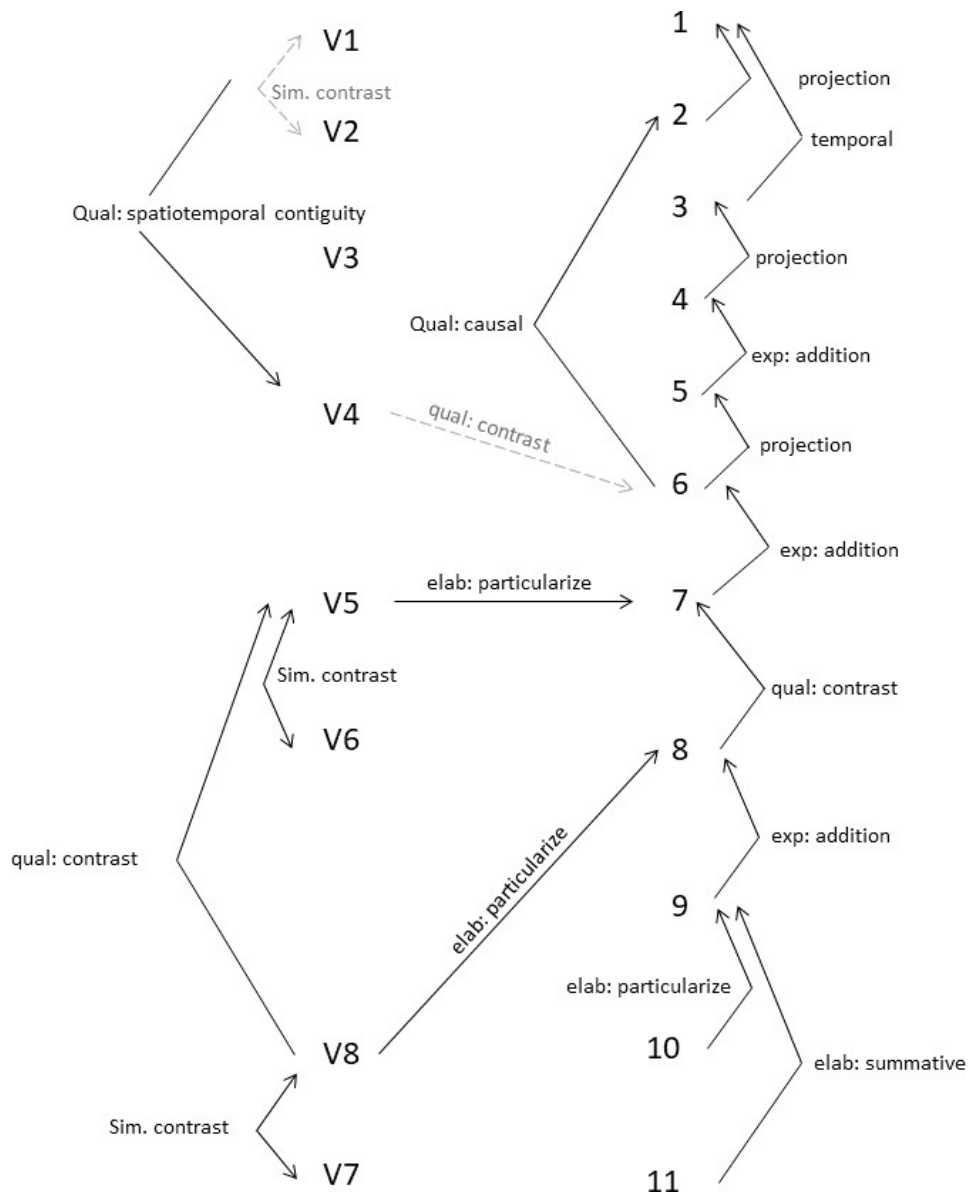**Figure 26: Logical conjunctions in Mamie's Dream, shots 1- 4**

7.4.1.2 *Congo VR*: Phase Analysis

Congo VR was analysed across shots C#16-21 encapsulating a phase that introduces the pygmies, their historic external oppressors, the consequences of colonial rule and ends with the introduction of postcolonial indigenous oppressors. The analysis is represented as reticula in Figure 27, p. 165.

Shot	CRU	Visual	CRU	Verbal
#16	V1	Forest	1	Far from the capital, the wounds of the past are still raw
			2	(=) Deep in the forest, pygmies live a life where little has changed since Belgian colonial times
#17	V2	Tribesmen	3	(+) This is Buyanga village.
	V3	BBC crew	4	(+) they're performing a good luck ritual before a forest hunt.
	V4	Village	5	(+) and we're going with them
#18	V5	Tribesmen	6	(+) They're looking for animals and useful plants along the way
	V6	Forest		
#19	V7	Man digging	7	(x) They've always hunted like this
	V8	Tribesmen	8	(x/+) but a century ago, their brutal Belgian rulers discovered something in the forest worth a fortune
	V9	Forest		
#20	V10	Pygmy elder	9	(=) Rubber
	V11	AL	10	(=) It's that white sap oozing out of the wild vines.
	V12	Tribesmen	11	(+) The pygmies were forced to search for them, and to collect it.
	V13	Forest		
#21	V14	Pygmy elder	12	(=) Our own sisters suffered at the hands of the colonizers collecting rubber
	V15	Tribesmen	13	(+) rubber was our heritage,
	V16	Forest	14	(+) and we did not benefit from it
			15	(= 12) The suffering started with our ancestors
			16	(+) and it continues into today
			17	(= 12) The white people, especially the Belgians seriously mistreated them.

Table 14: CRUs used in conjunctive phase analysis of Congo VR, shots 16-21

Verbal conjunctions span the three main types of expansion (addition, elaboration, and qualification/enhancement). Extensions of *addition*, both implicit and explicit, are found most frequent where new information is moving the sequence forward. Elaborations are used for locative purposes, like the visual CRUs where Alastair Leithead particularizes the forest as 'far from the capital' (CRU 1) to be *more specifically* 'deep in the forest' (2). Elaboration is also used to provide descriptions of 'something in the forest worth a fortune' (8), first by *presenting* it as 'rubber' (9) and then elaborating by *putting it another way*, 'that white sap…' (10). Some CRUs are slightly ambiguous such as 'they've always hunted like this' for the way it expands 'they're looking for plants and animals…'. This relationship can be an enhancing one, where the fact that

they've always hunted this way provides a circumstantial feature of *manner* (hunting traditionally), or it can be a temporalizing conjunction where 'always hunted' beds the activity in deep contiguous time. An adversative addition is used between CRUs 13-14 where 'rubber was our heritage' [but][30] 'we did not benefit from it'. Greater logical leaps are made across clauses where an initiating clause 'Our own sisters suffered at the hands of the colonizers collecting rubber' (12) is conjunctively related to two subsequent clauses: 'The suffering started with our ancestors' (15) and 'The white people, especially the Belgians seriously mistreated them' (17), each one elaborates the 'suffering' by presenting more specific information.

A macro-conjunction can be found spanning almost the entire phase, where CRU1 'the wounds of the past are still raw' is *explained* in clauses 15 and16. A feature of Alastair Leithead's speech across the text is his use of explicit conjunctions where they are not required (used to simulate interactivity, discussed in chapter 8 Tenor) such as his frequent use of 'and' in textual theme position.[31] Here 'but' is used similarly in this phase as an explicit conjunction between CRUs 7 and 8, where it is not immediately clear what is being compared. His use of 'but' must be interpreted instead in a more macro sense, as a means of injecting a macro-conjunction of *adversative addition* between the combined preceding clauses (the pygmies' indigenous culture) and what follows. In this way, 'a century ago, their brutal Belgian rulers discovered something in the forest worth a fortune' becomes a fulcrum around which the phase pivots into adversity.

The intra-visual conjunctions present largely as negative contrasts between visual CRUs (*similarities* between participants result in episodes that are themselves essentially CRUs). Notable contrasts include the BBC crew (V3) and the tribesmen (V2) where each form a salient episode, on either side of the spectator. Unlike the example given in *Mamie's Dream* (shots 3-4), this contrast does not appear to be part of a larger implicit logical meaning and is rather resolved cross-modally by Alastair Leithead when he states, 'and we're going with them'. Alastair Leithead's reappearance in C#20 creates another contrastive conjunction between himself and the pygmies that also does not seem to give rise to any larger logical meanings. His repeated appearance is at most an assumed *temporal succession* (*later that day*) and not significant beyond the pygmies' own successive reappearances.

Inter-shot visual conjunctions expand each other in several ways, by *particularizing* from general to specific entities, by *extending space and time*, and by *comparing* visual entities. Inter-visual

---

[30] 'And' is used in his speech which complicates the reading. I am considering this an implicit conjunction and an erroneous use of positive addition
[31] Alastair's contrived 'and' is heard in C#4, 17, 28, 29.

CRUs connect this phase spatiotemporally in a more obvious fashion than in Mamie's. One conjunctive thread is of spatiotemporal contiguous succession, using a conventional approach to establishing setting and then moving the spectator experientially from one aspect of the location to another. In the first instance, the village (V4) *particularizes* the opening forest CRU by 'zoning in' from the general vicinity to the exact location. Once the hunt has commenced, the visual forest settings *qualify* both space and time, where there is a different hunt location in V9 which is a different forest location to the previous shot and thus a later time also, presumed to be happening contiguously. The pygmy elder ends the phase in shots 20 and 21 in the same environs (V15-16) and as such only *time* is qualified ('later – in the same place'). In terms of comparison, the tribesmen form an inter-shot logical thread when they are found to be variously *similar* and *contrastive* with their successive appearances across shots. They are *similarly* standing and upbeat in V8 and 12, and then *contrastively* downbeat and seated in V15. This forms part of a larger logical syntagm involving the pygmy elder's verbiage, where he is temporally qualifying CRUs 15-16 ("the suffering started with our ancestors" "and continues into today") which forms a larger complex with the visual the tribesmen (V15) who are contrastively downcast, compare to previous shots and less willing to share their gaze in the unfolding events.

Cross-modal conjunctions in the phase were found as having elaborative and enhancing functions. Where deixis is used, the verbal CRUs verbally elaborate by *presenting* the visuals: 'This is Buyanga village' (4) and 'they're performing a good luck ritual' (5) can be considered as elaborating the tribesmen's setting and process (V2) where the latter is elaborating through specification (we cannot know it is a 'good luck' ritual from the visuals alone). Temporal conjunction is found across modes throughout the phase where 'before a forest hunt' and '...along the way' circumstantiates the tribesmen temporally in shots seventeen and eighteen. An implicit temporal conjunction can be derived from CRU 5 ('and we're going with them') and the BBC crew (V3) when made explicit through an implied 'and *when they do,* we're going with them'.

An ambiguous conjunction arises between CRU 11 "pygmies forced [...] collect it', and the concurrent V8 of the man cutting rubber from the vine. The CRUs are directly comparable involving the same process. The resulting conjunction can be read as positive or negative comparison '*similarly* this man is collecting...' or '*unlike* their ancestors, this man freely collects' (in other words, not forced). This naturally has repercussions on meaning regarding the status of the man and his actions.
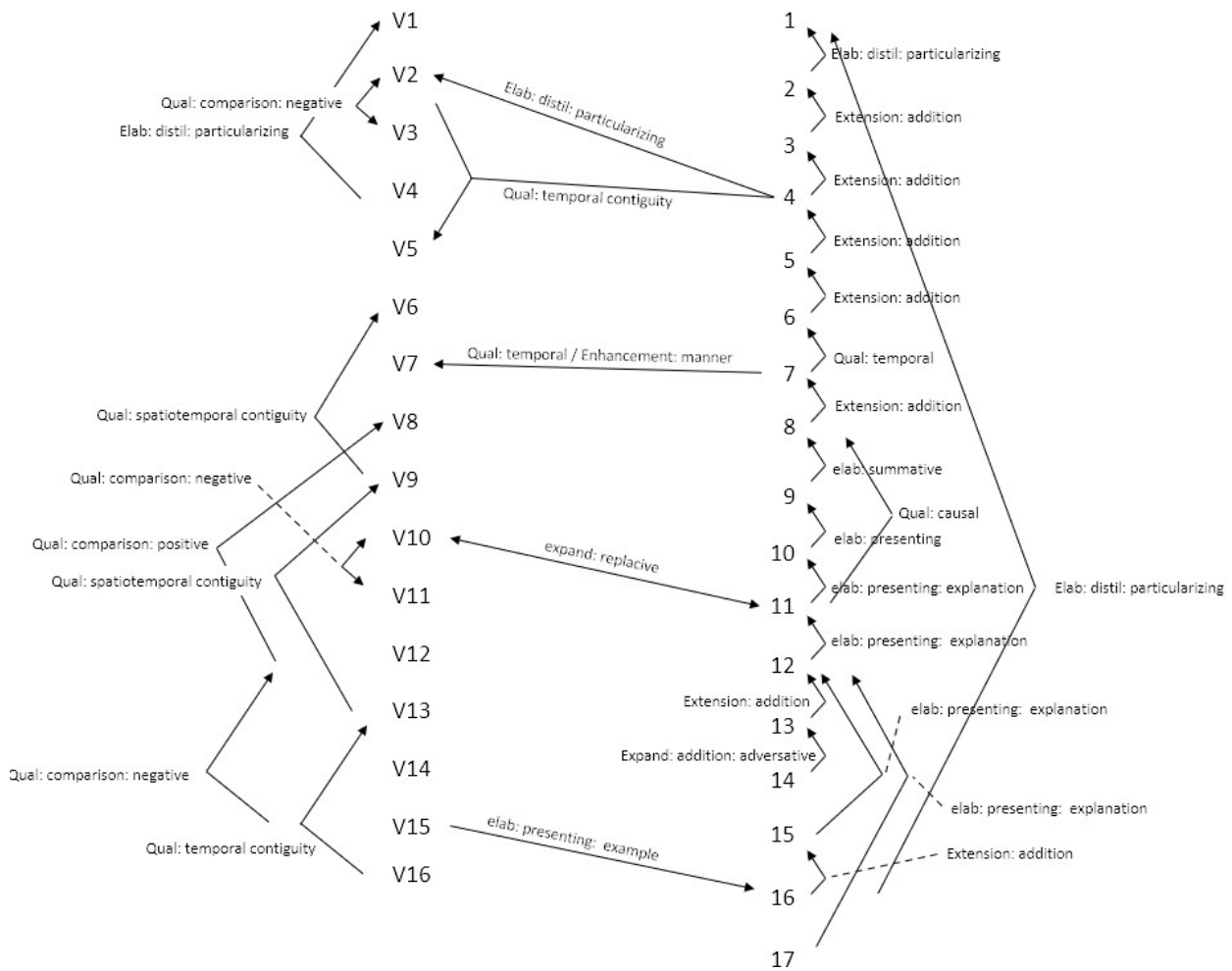
Figure 27: Logical conjunctions in Congo VR, shots 16-21

As general exercise, the conjunctive resources fit seamlessly into the analysis and were found to contribute to the semantics of the phases. CRUs were found to be generic enough to be applied flexibly to entities in the 360° image as well as in tandem with the lexicogrammar. As such, elaboration, comparison, qualification/enhancement were applied freely across CVR's modes realising complex conjunctive arrangements with meanings that included the split salience concept described in subsection 7.2, where rotation is motivated by the need to resolve intra-shot conjunctions. The analysis of *Mamie's Dream* was informative for the way conjunctions manifested in CVRs simultaneity and where rotation resolved contrasts required for the overall logical thread to manifest (boys-play freely as loved > girls play unknowing of their ominous fate).

Inter-shot comparisons and particularizations were invaluable in relating shot constituents across time and in some instances, *contrasts* did enable complex logical inferences at higher levels (as with Mamie described above). At the lower intra-shot level some of Martinec's conjunctions were less useful. Visual conjunctions did not always seem compatible with the explicit kinds of

165

conjunctions in the lexicogrammar. For example, *coordinating,* low-level conjunctions in the visuals (such as the teacher entering the room in M#6) seemed to offer little to the higher-level meanings in the voice-over.

The cross-modal conjunction of projection was deemed overly problematic and discontinued in the analysis. While on an intuitive level it is a feasible that where Mamie is seen internally focused and the spectator is hearing her subjective historic accounts, both can be fused in a single conjunction. It is difficult to support this however without linguistic support, such as by the possible use of deictic reference to synthesize the image and voice into a single event, for example, were she to lead with '*as I travel this river*, *my father's words come to mind'*.

7.4.2   Cohesion

The texts' cross-modal cohesive properties were discussed previously on the shot level in chapter 5, using contextualisation as the conceptual frame. This subsection extends that analysis across shots, using more low-level resources to establish the kinds of cohesive ties that 'glue' text together. The texts differ in their cohesive devices, reflecting their styles and requirements for 'localizing' the spectator in a continuous flow of activity. *Congo VR* threads together a series of locations with a degree of continuity at each location. This manifests as patterns of explicitly presented, specified, and reiterated entities maintaining a cohesive field of activity at each juncture. *Mamie's Dream* differs significantly. As a more disconnected sequence of locations and participants, bridging ties (meronyms, hyponyms, and synonyms) provide cohesion, both visually and linguistically. She does not explicitly present many of the critical participants and no-one is cross-modally identified either by verbal or graphical means. Also, many seemingly important participants are presumed, such as 'he' in 'he was older than me' (Mamie #7) who requires collocation ('he' + 'pregnant' + young girl), and a degree of extra-textual understanding that 'he' is her daughter's father.

7.4.2.1   *Mamie' Dream*: Cohesion Analysis

The phase from *Mamie's Dream* chosen for analysis is the same as used in the conjunctive analysis (shots 1-4) which comprises the following significant identity chains: *Mamie; Girls; Boys; Parents; FGM; River* and *Dream.* A synopsis is presented here of the more detailed analysis in Appendix H.2 Mamie's Dream Cohesion. Figure 28, below illustrates the way they cohere across the phase. *Mamie* is the strongest chain in the phase, where she is explicitly presented graphically in the opening titles (presumed, as we have not yet encountered her), verbally as personal pronoun 'I' and repeated visually in all but one shot. Her chain is cohesive by explicit means but intersects

with the *girl*'s chain through bridging ties. *Girls* forms a chain that stretches the length of the phase also (excluding the opening title), beginning with a series of verbal ties: the generic 'a girl' ties hyponymically with a more specific 'a little girl' (overlapping with Mamie chain), made generic again in 'girls' later specified again as 'their girls'. Finally, *Girls* becomes tied cross-modally in shot #4 with the immediately salient playing girls, who are also visually presumed in an antonymic, inter-shot bridging tie with the 'boys' chain, which terminated in shot#3.
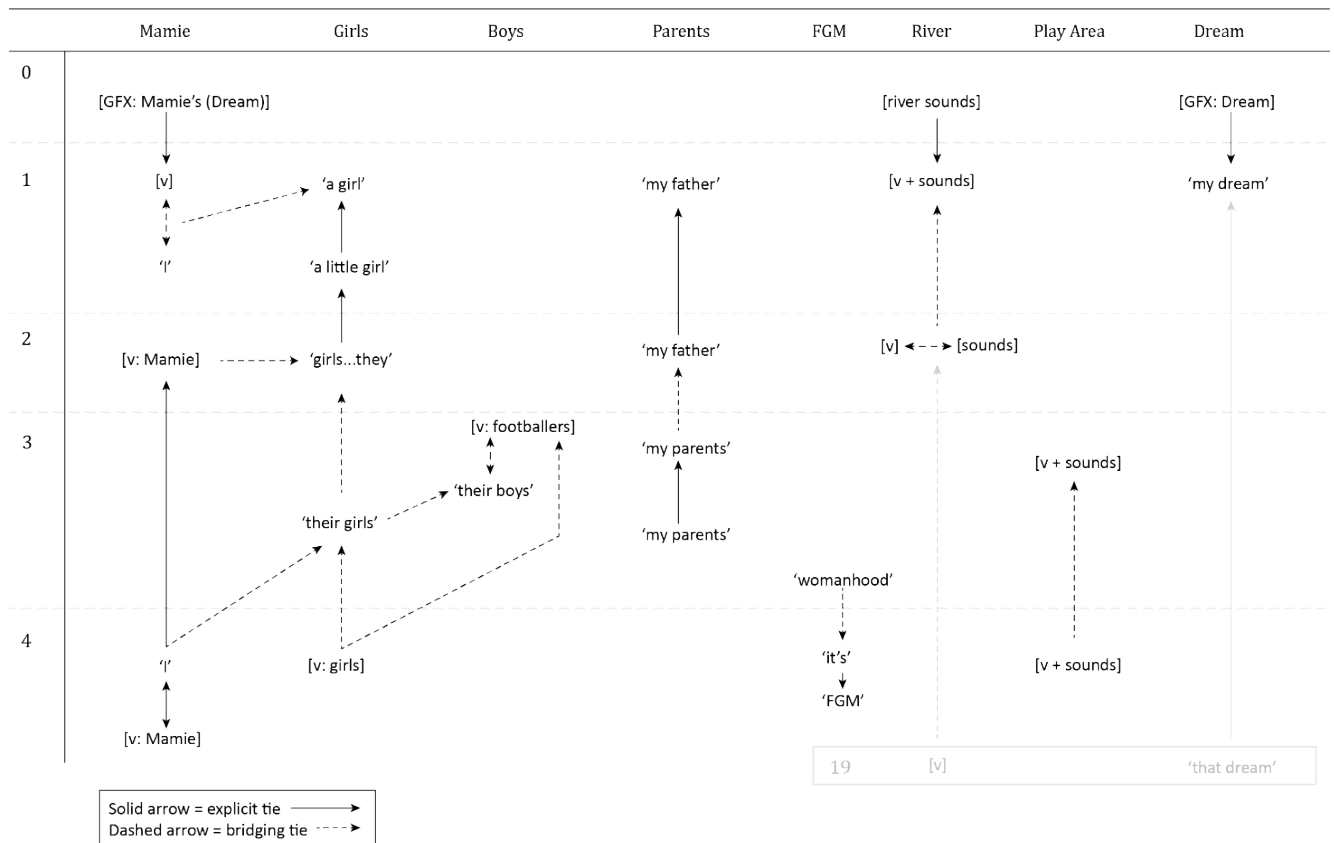
	Mamie	Girls	Boys	Parents	FGM	River	Play Area	Dream
0	[GFX: Mamie's (Dream)]					[river sounds]		[GFX: Dream]
1	[v] / 'I'	'a girl' / 'a little girl'		'my father'		[v + sounds]		'my dream'
2	[v: Mamie]	'girls…they'		'my father'		[v] ◄- -► [sounds]		
3		'their girls'	[v: footballers] / 'their boys'	'my parents' / 'my parents'			[v + sounds]	
4	'I' / [v: Mamie]	[v: girls]			'womanhood' / 'it's' / 'FGM'		[v + sounds]	
19		[v]						'that dream'

Solid arrow = explicit tie ⟶
Dashed arrow = bridging tie ---►

Figure 28: Cohesive identity chains in Mamie's Dream shots 0 (titles)-4

Boys is a short chain with just one cross-modal tie, '(loved) their boys' and the visible footballing boys. It would be possible to include both boys and girls in a more general chain of 'children' or 'boys and girls' but it is more analytically useful here to juxtapose the two chains where 'girls' has clear primacy and 'boys' when used as a shorter chain becomes a circumstantial feature of the phase.

The *Parents* chain includes a prominent *father* who could arguably have his own chain across the text, but in this phase the units morph between 'my father' (presented twice), 'my parents', 'my father and mother' and 'many families in Sierra Leone', and as such are cohesive as a bridging

chain of instances of 'parents' (as reiteration, meronym, and hyponym respectively). Here the chain moves from specific to generic as befitting Mamie's move from narrative to exposition. The *FGM* chain begins a reference chain in shot 3 where a cataphoric synonym of 'womanhood' obliquely establishes the presented 'FGM' in shot 4 (preceded by two more cataphoric references ('it's)). The length of the cataphoric segment of this chain reflects the suspenseful and impactful nature of her utterance.

Finally, the river presents a novel example of synonym, with the verbal clause 'just get married and *move away*' where 'away' and 'move' are both qualities of the river flowing toward the horizon. Zooming out to the macro-text, the river can be granted additional cohesive importance as it coalesces with her *dream* chain. As the first phase of the film, certain entities are introduced that will not be recovered until much later in the film such as 'my dream' (shot #1) which will not be reiterated until the final scene in the film ('that dream I had'). Connected with the dream chain is the 'river chain' which is concurrent with her dream chain (shots 1 and 19).

7.4.2.2   *Congo VR*: Cohesion Analysis

Congo VR (shots 16-21) presented the following chains: *location; pygmies; pygmy elder; man digging; ritual-hunt; rubber; colonizers; BBC crew; and suffering.* These chains are represented graphically in Appendix H.3: Identity Chains from Congo VR, Shots 16-21. As with *Mamie's dream*, a synopsis is presented here of the more detailed analysis in the appendices

Pygmies are represented in each shot visually, in the lexis, and in the diegetic audio. The pygmies are presumed initially in shot 16 through a generic mention of 'pygmies (live a life)' and the diegetic sound of men chanting. The chanting (heard only) forms a multimodal cataphoric tie or 'audio prelude' (Tseng, 2013) with the following shot, where pygmies are presented cross-modally (although never specified thereafter verbally by Alastair Leithead as *these* pygmies). The pattern thereafter is a series of ties narrowing from generic 'pygmies' to pronominals 'they', them', they', localized by their use as deictic references to the visualized tribesmen ('they're performing...', 'they're looking...'). As the phase develops, and the pygmy elder takes the floor, lexical ties of *hyponym* open the chain up to a wider contextual frame ('*the pygmies',
'our own sisters', 'our ancestors'*). This becomes coextensive in shot 21 with another chain, *suffering*. As the pygmies are historicized, the visual tribesmen, no longer in cross-modal parallel processes become instead hyponymic with the verbal chain (as *pygmies*, as *fellow brothers*, as *descendants*).

In this chain, the group of tribesmen are repeated across shots at roughly the same level

of generality, maintaining a stable 'pygmies' chain that allows other reference chains to emerge. For example, the *pygmy elder* (Ilungama Ayanda) is presented multimodally in shot 21 (visually, through diegetic speech and named in a graphic title) prior to which he can be considered part of the pygmy chain. His gradual specification in shot 20 extracts him from the pygmy chain but he remains ambiguous as to his presumed/presented status (he is oriented away from the camera). He is in fact also visible in shot 19 and can be considered to begin his chain as a cataphoric tie of meronym with the visual tribesmen but has no salience in the shot making it difficult to assign him a presumed status. Prior to his emerging, another participant *man digging* is given salience, but without any specification, prelude, or significant appearance thereafter. His chain does not develop and is as such somewhat ambiguous and incohesive.

Other chains such as *location*, *suffering*, *time*, *colonizers*, and *BBC* span the phase but not with the same frequency of ties as the pygmies. For example, the *location* chain begins with Alastair Leithead using a lexical tie of synonymy between 'far' and 'deep' (or possibly co-hyponyms both as metrics of *distance*) to orient the spectator, forming a prelude to the subsequent presenting of *'Buyanga village'* in shot 17. From there, the ties are less explicit and apart from a very vague 'along the way', the visible forest backdrop is the only remaining cohesive tie, as co-meronym with other forest locations and as repetition where two adjoining shots share the same location (e.g., #20, 21). The chain s*uffering* begins with 'wounds of the past' but does not have a tie until the last shot in the sequence where 'suffered', 'suffering', and 'mistreated' create ties of collocation (wounds-suffering-mistreated). 'Wounds of the past' is ambiguous in its status but is more likely interpreted as a cataphoric tie, requiring the last part of the phase for its meanings to be realised.

Finally, reference chains were found to morph as the phase progressed adding a degree of ambiguity, possibly disrupting the phase's cohesion. For example, *ritual-hunt* begins in shot 17 where the tribesmen are seen and heard as well as verbally referenced in hunt related activities. The chain coheres thereafter in the lexis through ties of collocation of *hunt-animal-hunt* ('…before a forest *hunt'* > looking for plants and *animals'* > *'hunted* like this') as well as synonym (*'looking* for plants and animals' > *'discovered* something'). This sub-phase of hunting coheres strongly but terminates abruptly, replaced by a *rubber* chain in terms of the *searching*, *looking,* and *collecting*. This morphing of reference chains occurs in the clause 'discovered something in the forest worth a fortune' which is both anaphoric (ties back to the hunting through synonym of 'looking' and 'hunted') and cataphoric where the 'something' is presented cross-modally in the next shot ('rubber'). Whether this is by choice or happenstance, the result is that the hunt chain continues but informationally it becomes concurrent with the rubber chain and possibly incohesive with

that which preceded it.

When both films are considered, their cohesion differed significantly reflecting styles and requirements in representing their subject matter in the phases. *Congo VR* constructed spatiotemporal continuity through the reiteration of its subjects between shots, where Mamie was the only explicit continuous tie made through the phase. As befitting an autobiography, she formed the cohesive backbone of the phase where others were represented as a constellation of characters connected through bridging ties of meronym, hyponym, and synonym. Where the repeated *pygmies* constructed a spatiotemporally contiguous sequence, Mamie's participants remain more general pointing to kind of field of activity she is construing that is not in either her or the spectator's purview. Bridging ties also allow Mamie to weave a subtle exposition regarding FGM, as the culmination of a series of cataphoric presuming references tying cultural attitudes to 'womanhood' to the horrors of FGM. Other more subtle cross-modal ties, such as the synonymous tie between the river and her father's projected '…just get married and move away', are an opportunity to consider the connection between the spectator's rotational act to underlying semantics. By rotating from Mamie (the salient entity in the shot) to survey the scene, the connection between the verbal and visual is augmented and rotation 'coheres' text. A possible issue with this however is whether the tie is *required* for the text to cohere. In this light, it might be more suitable to consider the tie as an enhancing conjunction, such as '*like* the river…'.

The analysis of *Congo VR* showed continuity, as a series of explicitly presented entities tying through repetition which carries risks, where chains once established are expected to be realised explicitly. In this light, the kinds of continuity-based cohesive devices described by Tseng can be difficult to maintain across 360° shots. For example, Tseng's 'physical preludes', used in her analysis of Memento (2013, p.84) show how close-ups form cataphoric ties. In her analysis of frame-based films, an entity is presented gradually where a close-up of a hand, for example cataphorically references the character to whom it belongs, presented thereafter in full. The CVR camera does not facilitate close-ups as cropping effects, and visible preludes must be achieved by different means. The pygmy elder is of interest here: In shot 20, he is cataphorically referencing his own multimodal presentation in shot 21. The prelude is quite different to Tseng's as he has already been presented through his immediate salience (#20) as he hacks the rubber from the vine, surrounded by onlookers. Added to this, he is in fact visible in shot 19 but with no salience, in which case he is technically a prelude here also but more likely remaining in the 'pygmy' chain. His ambiguity, determined in part by the CVR image modality is not a major problem for understanding the scene as the spectator can resolve his emergence into his own

170

identity chain. What is of significance here is that without a frame, entities in CVR *must* interact with other reference chains before establishing their own.

Another shot in *Congo VR* is perhaps more problematic. In section Framing7.1 I discussed Junior in shot 13 from the point of view of erroneous framing and salience, where Junior's female on-stage companion becomes salient but remains generic. When considered temporally, this is a matter of cohesion regarding the identity portion of the reference chains that manifest in this phase. Junior is explicitly presented cross-modally, as visually salient and addressing the spectator. The woman's status however is awkwardly somewhere 'between' him, and the people below; not fully generic but not fully specified. Her visual salience, augmented by the gaze interaction with Junior marks her out as having a chain of her own, but one that essentially dissolves in the next shot where her salience is reduced, as she merges with the other generic participants in the yard (this effect is found also in Mamie #15-16 but in that case the girl's chain is re-established in #17). For context, a distinction can be made here between the woman in Congo #13 and other 'fully' generic participants such as the highly salient but unspecified woman in Congo #10, where the latter has no other resources used to single her out and importantly, is not repeated through other shots. I would take the position here that the issue presented in shot #13, although slight, points to the consequence of the CVR's frameless modality regarding cohesion.

7.5   Concluding Remarks

This chapter is information-focused and as such primarily addresses the first and third research question, regarding the configuration of discourse in CRV and the adaptation of SFMDA systems to facilitate CVRs information structures. Secondary to this, there are inferences for producers of nonfiction CVR texts, which will be discussed here in brief.

Principals of framing, salience and information value were adapted for CVRs simultaneous space. The critical aspect of framing in CVR is that the spectator is part of a continuous 360° space and will by necessity interact with the framing devices used. 'Stage' and 'container' frames appear to be natural and common-sense ways of describing the connecting and disconnecting of things relative to the activated ground-plane. CVR renders these objects as experiential things also, and their framing is in this way tautologous: a boat is a boat, regardless of whether a producer also wishes it to be a frame. This does not reduce the capacity of recognisable objects to frame, it merely highlights the importance to recognise them as frames in the production of nonfiction texts. 'Windows' were presented as a way of organizing information into a

visible/imminent, which are useful as junctures between what is in the spectator's purview and what has meaning yet is occluded from view. The dichotomous aspect of this kind of frame ('in or out') suggests perhaps that it could as easily be a component of a particular kind of *information value* in CVR, akin to *visible-imminent*. Salience was described as a fundamentally transformed concept in CVR, where field-of-view is required to enable gradual and split salience across entities.

Framing and salience can be described as co-extensive with cohesive systems. Returning to C#13, an overlap is found between the *framing* connecting Junior and his acquaintance, lending her importance, over the generic participants below, which coincides with her proximity, size, and overall *salience* in the shot. This marks her out as a quasi-*presented*/*presumed* entity, who is incohesive. The converse exists in M#16, where the girls is *not framed*, *not salient*, and *incohesive*. This points to the structural issue with CVR for producers, where the meanings at risk in the previous chapters are found in information-structures. All entities in the 360° image are interacting with a framing device, and all entities are thus interacting on some level with a reference chain.

Analysis of reading path was somewhat valuable but limited on a deeper level. The value lay in uncovering the semiotic choices made by the producers endeavouring to lead the spectator thought the text. It is also valuable to gather and connect the disparate resources that can lead attention around the image. The limitations lie in the idealising of discrete reading paths, where, given the length of the shots and the rotational freedom in CVR, these reading paths are far from guaranteed. Also, 'weakened' or 'broken' paths can be quickly compensated for, by the spectator. While continuity is fundamental to analysing reading paths, discussing the effects of continuity also presupposes a mode of filmic engagement based on classical frame-based naturalistic editing. For example, it is tempting to conceptualise a disrupted RP as a *jump cut* that brings the spectator out of the experience, but this would assume they were 'in' an experience of continuity in the first place, in what is a highly reflexive medium. Finally, there was an issue whereby the intra-shot RPs did not necessarily join salient entities. For example, Mamie's gaze to the horizon in M#2 sets up an RP akin to *POI* (Mamie) > *context* (river) which connects across the edit with the salient footballers. It is difficult to rationalise this, excepting for the fact that both shots match by virtue of each having a dichotomous mise-en-scene, allowing binary choices (Mamie/river || footballers/watchers). Studying how CVR films *cohere* appeared more fruitful to a textual analysis such as this, as it addresses the threads of meaning that exist in the text that can be realised by the spectator.

8. The Texts as Situations of Nonfiction Genre

This chapter addresses the research questions at a higher conceptual level than the previous chapters. The texts are described here as instances of semiotic technology and as situations of nonfiction discourse. Through the register variables of MODE, field, and tenor I will discuss the texts in the context of their technical affordances, adding a semantic dimension to immersion, presence, and co-presence. This in turn frames the discussion on the role and effects of language in the texts as a means of negotiating the limits and affordances of the immersive modality.

As laid out in chapter 3, section 3.2 (The stratified modelling of language), register interfaces 'above' with genre, and as such the texts are considered here for the ways that lower-level semantics realise nonfiction genre, through socio-semiotic activities. As a frame, I will use Martin and Rose's *staged, goal-oriented social processes* (Martin and Rose, 2008) to describe how the texts share certain nonfiction 'family resemblances' (Plantinga, 1997). Both texts are essentially didactic, taking the form of Reports,[32] but they each mix other generic elements such as story narratives, historic recounts, as well as novel genres such as guided tours.

I have alluded in previous chapters to the semiotic autonomy given to the spectator as they construct text, and the kinds of practices required by the producers of CVR texts. As a higher-level analysis, this chapter will allow for a more comprehensive and holistic account of CVR in this regard. When concluding this chapter I will expand and connect on those themes, drawing out the underlying semiotic, ethical, and ideological aspects of the CVR as nonfiction text. In doing so, I will reintroduce performativity and other concepts discussed in the literature review, to add to the discussion.

8.1   Mode

The role that spoken language plays differs across the two texts resulting in different kinds of situated discourse. This is a fundamental aspect of MODE with consequences for the other register variables. The sharp contrast in the roles that language takes is illustrated in the following extracts:

1) 'At the entrance on your left, there were once Picasso's painting' (Osambia-Kpwata Fyfy C#26)

2) 'My father said educating a girl is a waste of money' (Mamie M#1)

---

[32] There are two uses of this word: as a *report* genre (as noun) and *reporting* as semiotic activity (as verb). For this reason, I will capitalize the former all instances of Report as genre.

In the first extract, language is used as if the speaker and listener are co-located in a contemporaneous exchange, where context of situation appears to resemble an actual conversational situation. The second extract features speech that does not appear to be tied to a specific situation as it could be delivered in any mediated or unmediated context (in person, on the phone, as a recorded video message, for example). The first extract makes mention of the listeners embodiment in the situation, where 'your left' suggests both an awareness of the spectators' orientation (looking forward) and the imminence of their rotation (looking left). The second example makes no such situational claims and directs the spectator imaginatively to an unseen 'father' and his speech act, constituted entirely through her voice-over.

Even though it is not actually the case that Osambia-Kpwata Fyfy's speech acts involve a contemporaneous interlocutor, this part of the text is carefully designed to foster the illusion that her speech is occurring in the spectator's here and now. In a systemic functional sense, this can be conceptualised as the text generating a multimodal *virtual context of situation*. While all texts do this to a degree, CVR's unique kind of virtualised situation is enabled by the 360° image's verisimilitude and place illusion, which facilitates the sense of *situation* over *image*. The significance of this goes beyond the psychological effect of presence in a situation, as it is the situation of *discourse* that is virtualised. In this sense, SFL's mediating of the contextual stratum allows register to be considered for the kinds of linguistic devices used in the texts, where language is either *embedded* in the virtual contextual field (as in example 1 above) or *constituting* a different field (as in example 2).

This is both a notable aspect of CVR's technological capacities and a necessary framing device for analysing CVR's register. Given the kinds of language choices shown above, there is a clear need to consider the CVR image as a kind of intermediary context of situation *as-if* it were real. It must be stressed that this is a necessary artifice on my part for the purposes of describing CVR as a semiotic technology, where technically nothing viewed in the headset constitutes an actual immediate situation or field of reality. Osambia-Kpwata Fyfy's language-use is no different to Mamie's reflections on her father, both being subject matter consigned to the past. Also, it is clearly not the same for a disembodied voice-over to be part of an actual immediate situation, but it is possible to critique the *use* of a voice-over that purports to be part of the immediate situation. I am using this conceptual move, as a necessary frame for discussions in this chapter as it enables deeper analysis regarding choices in language and a characterisation of CVR's high-level semiotics. For clarity, I will prefix aspects of context with 'virtual', where necessary such as, 'virtual situation' and 'virtual field of activity'.

8.1.1   Channel and bandwidth

As a communication technology, CVR's novelty lies in its sensory mix. Comprising aural and visual sensory channels, CVR is much like the framed linear audio-visual forms it remediates. The 360° image distinguishes itself however through the augmenting sense of *presence*. As a purely psychological phenomenon, presence must be formulated in more semiotic terms, to describe "the potential for different combinations of semiotic systems" (Matthiessen, 2009, p.25). The technical aspects of immersion that facilitate presence are more apt in this regard, such as the rotating perspective (CVR's 'three degrees of freedom') and the erasure of the HMD. Rotation and erasure combined facilitate an embodied reaction to the environment through orientation and movement, which extend the sense to *exteroception* and *proprioception*, respectively (Sutherland, 2012a, p.47; Bertrand et al., 2018)). Sense, in this regard becomes *sensorimotor* at the lowest level, and when combined with the aural and visual senses, can be considered as 'sense making' within the environment (Passmore et al., 2017).

CVR's channel becomes informational as bandwidth, as the 360° image configures the flow of sensory information to, and around the spectator. Spatiotemporally, CVR presents a paradoxical schism of perceptual immediacy and asynchronicity. The indexical mode of 360° image capture and immersive sound results in highly detailed place illusion, which is encoded and played back as linear video. In this regard, CVR is an information rich channel but with contact that is essentially one-way given that the spectator cannot modify the materiality of the text in their interactions.[33] This notwithstanding, place illusion and sensorimotor engagement points to CVR's novel bandwidth. Spatially, the rotating spectator's exteroceptive senses and object permanence leads to a sense of a 'fixed' space and contact with information that may or may not be currently in their field-of-view. Temporally, the rotating movement is reflexively the spectator's embodied time, and with that comes a sense of time that is independent of, but concurrent with, the timeframe of the asynchronous text. This would appear to be CVR's defining characteristics expressed as bandwidth, where spatiotemporal contact is sensorimotor and contextualising which is qualitatively different to that afforded by synchronous audio-visual communication forms. 'Immediacy' can be framed as the spectator's reflexive temporality that creates a paradoxical 'contemporaneous-asynchronous' experience.

To articulate CVR's bandwidth further I will describe the way it manifests in the texts as *co-presence*, a defining feature of the form when a filmed subject is involved. Co-presence can be

---

[33] At a very low level, the spectator does modify the text as each rotation technically renders a grid of pixels from all available pixels. On the more meaningful level used here, the spectator cannot alter the contents of the image, or generate new content.

considered as the product of CVR's channel and bandwidth, of being reflexively aware of being 'there' yet unable to interact with the representations of other participants in any significant way. Co-presence is assumed in both texts, particularly in *Congo VR*'s direct address, which strives to simulate a degree of prototypically high-bandwidth conversational encounters. Conversation is held as the 'gold standard', prototypical information exchange, occurring face-to-face in a singular space and time, through all sensory modes. Although real-time feedback is missing in CVR, contact is reflected in the contextual information afforded the spectator. Assuming rotation occurs, there is the increased availability of a speaker's paralinguistic cues, necessarily available in CVR (discussed in section 5.1 as 'omnipresent low-level processes'), which would otherwise have required framing for inclusion (as 'cut-in' shots). Also, there is the necessary inclusion of other participants and aspects of the mise-en-scene who add contextual, extra-linguistic information to the speaker's utterances. In this fashion CVR provides an 'always-on' contextualisation for whoever has the stage, always adding information by default and always allowing the spectator to process the speaker's utterances as the product of the overall virtual situation. This is evident in Ilungama Ayanda's intense gesturing (C#21) and forceful utterances which are intensified by the other tribesmen who are impacted by what he is saying. It is also evident in C#14 where Princess Mamicho's enthusiastic gesturing to the camera take on an ironic quality when taken in the context of the others in the scene, who are not part of the performance. While all this context could be included in a framed continuity edit, it is the uniquely sensorimotor, sense making aspect of CVR's bandwidth that enables the flow of information.

On a textual level, the specific modal ensemble will have a configuring effect on bandwidth. In this regard, modes, such as diegetic voice, voice-over and visuals are *textual channels* that mix and match in the text, addressing the senses in particular ways. While the texts address the spectator sensorially with a voice and an image of the speaker, it is not the same to hear and see Alastair Leithead speaking in the diegesis (C#4), as it is to see him looking out the window while hearing him in the voice-over (C#8). Neither is it the same to hear Osambia-Kpwata Fyfy's diegetic voice (#26) carry over into the following shot, as voice over. All these variations modify bandwidth, in ways that can be quantified through the mode variable *medium,* and the roles given to language.

8.1.2   Medium

All CVR representations have the technical bandwidth to appear somewhat contemporaneous. This is used to greater and lesser degrees in the texts to make situations seem *spontaneous*, as if constructed in an ad hoc situational fashion. Naturally, the texts, being asynchronous, do not

have the capacity for actual spontaneity yet their medium includes speakers using language in a multitude of ways, including dialogically, while using their speech to further construct the simulated situation at hand. The choices of language employed in the texts can be described as varying between *spoken* (i.e., dialogic, context dependent, dynamic, and spontaneous) or *written* mode (monologic, context independent, synoptic and in 'final draft' mode) (cf. Eggins, 2004, p.93).

C#4 is an example of medium and bandwidth coinciding as relatively natural conversational bandwidth. Here, Alastair Leithead is seen and heard in the diegetic track, using speech that further supports dialogic MODE, containing markers of spontaneous and interactive speech construction.[34] By referencing his surroundings ('this mighty river') he references situational context, indicative of a co-spatial encounter. He also injects an unnecessary continuative "*and* the team and I", suggestive of urgent construction. Only Princess Mamicho's exchange surpasses his with a tagged clause ("you see what I mean") and highly personalised spectator ("I am so glad you have come to visit"). While diegetic speakers can maximise bandwidth with dialogic speech, a more strategic use of medium is found where dialogic language is used in the voice-over to bring the speaker and spectator into a quasi-dialogic contact with the visuals. C#20 pairs dialogic speech with non-interactive visuals. Here, Ilungama Ayanda, Alastair Leithead and the tribesmen, are accompanied by dialogic medium in the voice-over.

> "Rubber. It's that white sap oozing out of the wild vines. The pygmies were forced to search for them, and to collect it."

Here, speech is dialogic in its sparse lexis and deictic reference ('*that* white sap'), suggesting context-dependence. Like Osambia-Kpwata Fyfy in C#26, Alastair Leithead is 'playing' with MODE as a way of engineering a dialogic space between himself and the spectator, without recourse to his diegetic voice. Where conversational bandwidth is not a necessary aspect of a shot, the speakers have free reign to use monologic 'written' mode. Alastair Leithead's lexis is dense[35] and complex in places reflecting a carefully drafted, polished and synoptic writing style. His many long noun phrases are unlikely to be found in typical conversation such as in C#3 where he describes The DRC as "...a beautiful but troubled place", packaging an otherwise lengthy series of logical conjunctions into a single nominalized phrase. Similarly, his use of metaphor (e.g., "the wounds of the past are still raw" C#16) is monologic and typically found only in written prose and scripted oration.

---

[34] Alastair Leithead's lexis comes down to a spoken density (44.4%).
[35] Alastair Leithead's voice-over frequently inhabits the ranges of lexical density associated with exposition. (50%-60%) and technical writing (60%-70%) see appendix G.

Other speakers in *Congo VR* inhabit a middle-ground of sorts mixing spoken and written medium. Osambia-Kpwata Fyfy and Ilungama Ayanda exhibit roughly the same lexical density (approx. 50%) but differ along the other dimensions. She personalises both the spectator and herself, whereas he does neither; she tags her clauses with '…I told you', lending her speech a degree of spontaneity, whereas his delivery is straightforward as a series of declarative statements. Osambia-Kpwata Fyfy's lexical MODE is most notable for the way she directs the spectator's attention, signalling a rare use of context dependency and actual spontaneous textual co-construction (should the spectator in fact heed her prompt and look 'to [their] left').

Mamie's Dream does not employ the explicit markers of spontaneous construction used in *Congo VR*. She is the sole voice-over eschewing direct address and any direct acknowledgement of the spectator or her immediate context. As such, medium is manipulated in more subtle ways and with fewer resources. Cues are found in her lexical density which fluctuates across the text indicating a subtle shift between *speaking at*, and *speaking to*, the spectator (see Appendix G: Lexical Density Analysis, p. 2 for charts illustrating her variations in lexical density). In exposition mode, Mamie reaches Alastair Leithead's high levels of density[36] losing any vestiges of dialogic exchange in favour of heavily constructed, polished, and synoptic speech:

> "My **parents loved** their **boys**. Like **many families here** in **Sierra Leone**, my **parents wanted** their **girls** to **join** the **Bundu society** and **enter** into **womanhood**"

In the example above, her 'parents' (twice), 'boys', 'many families', 'Sierra Leone', 'girls' and Bundu Society are all invoked in her proposition, with *love, being-like, wanting, joining, becoming women* all collated into two clause complexes. Condensed constructions such as this are not typical of conversation. Where this mode of speech is 'speaking at' an audience, it is feasible in other shots to interpret her words as if spoken to an interlocutor. In these instances, her speech tends to span singular clauses or short clause-complexes, requiring more grammatical (i.e., non-content) words for coherence, such as in the example below. When there is a high volume of these short clauses, it also enables empty spaces between utterances to simulate interactivity through the effect of her letting her statements 'sink in'.

> "I **refused**. I was **scared** I might **die**. I was **twelve** at that time" (M#5, 35%)

In between these extremes, Mamie's lexis inhabits the *fiction writing* range of lexical density[37].

---

[36] Mamie's lexical density reaches 'technical density' (60-70%) in shots 3, 4; 'exposition density' (50-60%) in shot 16.

[37] Fiction writing density is between 45-55% found in M#1, 6, 10, 12, 13, 50.

In this mode she is using content words to push the story along at critical narrative junctures.

> "I **heard** it's **painful**: they **blindfold** you and **cut** your **clitoris**. It's **called FGM**"
> (M#4, 50%)

> "When my **daughter went** to **school**, I **decided** to **volunteer** there. **One day** the day
> the **headmaster asked** me if I **wanted** to **join** a **learning assistance training
> program**. He didn't have to **ask** me **twice**" (M#10, 50%)

Compared with the contrivances used by Mamie, Alastair Leithead and Congo's social actors, the generic speakers in the texts are also notable their voices in the film, and for the ways channel and bandwidth 'naturally' coincide with medium. These voices occur most notably in *Congo VR* such as the woman's loud vocative towards an unseen listener in C#10, and in *Mamie's Dream*, where an unnamed and unrepeated teacher's voices are heard. These voices are interesting as they can be considered for having register of a different order to the specified participants in the films. Although her speech can only be retrieved in part, the woman at the train station is enacting an actual spontaneous, on the fly, dynamic construction, with contextual dependency '*Francise*, [undecipherable] *est la!*' ('Francise…is there!'). There is no textual strategy in play here and medium naturally follows channel and bandwidth, where aural, visual, and potentially sensorimotor bandwidth are in synchrony. The teachers' utterances in *Mamie's Dream* are similar if slightly more complex, as they are already enacting a *teaching* register.

8.1.3   Experiential Distance

The woman in C#10 who is shouting to the unseen '*Francise*' is using language to draw the attention of someone to a place that is in their immediate vicinity, and as such her language is embedded in the social process at hand. Conversely, Mamie's remembrances at the start of her film, are social processes in themselves that function to constitute experience (Eggins, 2004, p.91). As action-based and reflection-based respectively, they inhabit the poles of an *action-reflection* continuum of experiential distance[38] (Martin, 1984; reproduced in Eggins, 2004). These examples show extreme variation in the experiential distance (ED) between language and experience. Other instances from the texts fall along the continuum and can be described for the way they exhibit and at times 'play with' ED to maximise the sense that language and social processes are fused in an immediate multimodal experience.

Discussing ED as the relative distance between language and 'social processes' requires that

---

[38] Eggins uses the analogies of playing a game > commenting on a match > recounting experience > constructing experience (as fiction/nonfiction).

these processes are considered as those occurring in the 360° image, as the *virtual context of situation* discussed at the beginning of this chapter. By bracketing out the fact that all textual representations are in fact *constitutive* of experience, ED can be discussed in a fulsome manner. For example, the use of deixis in *Congo VR embeds* language explicitly within social processes such as Alastair Leithead's uses of deictic reference in "*That* cross now filled with water", where he closes the semiotic distance between speaker, subject, and spectator. Further along the ED continuum, *Congo VR* includes language that is more of an *accompaniment*, 'commenting' on unfolding social processes. These were reflected in the film's many co-contextualizing structures described in section 5.3.1 where speech elaborates visual processes, such as in Alastair Leithead's "They're looking for animals and useful plants along the way", commenting on the visible pygmy hunt (C#18). These kinds of parallels and ED are few in *Mamie's Dream*, barring moments such as "The students that I teach call me teacher" (M#12). Mamie's more complex intersemiosis presents an ambiguous pattern of ED. For example, M#3 sets up a cross-modal cohesive tie of hyponym between the boys in "My parents loved their boys" and the visual boys playing football. Where such bridging ties are used, it is plausible to consider ED as shading between *reporting* on and *constituting* experience.

Tense also alters ED, where language construes historic processes, or processes that have yet to happen. Language in the past tense, that has no direct parallel in the visuals is by its nature reflecting on experience. In *Congo VR*, Ilungama Ayanda reflects on his people's colonial past; Alastair Leithead and Osambia-Kpwata Fyfy reflect on historic social processes pertaining to her father's palace. As an autobiographical text presented largely as reportage, Mamie presents an interesting use of ED.

> "I heard it's painful: they blindfold you and cut your clitoris. It's called FGM" (M#4)

Here distance is two-fold: by recounting her experience of her hearing, she allows for ED indicative of both *reporting on* and *constructing* experience. Future tense also establishes ED. In the following examples, Eggins' *sports commentating* analogy is applicable where Alastair Leithead is providing linguistic support for the ensuing journey and Mamie, for her future independence.

> "And the team and I will be taking you on a great adventure up this magnificent River we'll be exploring its history, its riches, its poverty, and its future" (C#4)

> "Soon I will qualify as a teacher and won't have to depend on anyone for money" (M#15)

In both cases, experience is both contemporaneous and imminent: the 'team and I' and 'this magnificent river' are all visually present and available as part of an unfolding social processes,

but the 'great adventure' is not contemporaneous in the same sense as the hunting pygmies in C#18. While language might be *commentating* on imminent experience in these examples, is also *constructing* those experiences.

Experiential distance can be considered as a dynamic resource. Alastair Leithead is notable for the way he 'expands' and 'contracts' ED as a means of reflecting on expository aspects of the text, while anchoring his exposition in the immediate situation. When Alastair Leithead is describing complex DRC-related matters, he expands between language that is experientially close (embedded in experience) and distant (constructing experience), often in the same clause-complex. This is exemplified in the extract below which I have divided into relevant junctures on the action-reflection ED continuum.

1. Language accompanying experience: *"The sapeurs display the energy and brash confidence of the capital."*
2. Language reporting experience: *"Two-thirds of this country are aged under twenty-five".*
3. Language constructing experience: *Young and optimistic, they're desperate to burst free of the country's troubled history*

(C#15)

AL's expansion and contraction of ED signals his desire to structure his use of language to both support the experiential aspect of immersion and presence, while using this immediacy to anchor his more rhetorical constructs. When CVR's technological aspect of bandwidth is factored in, his manipulation of ED takes on additional resonance. In this light, Alastair Leithead's rhetorical expansion becomes a visible and embodied expansion also, extending the spectator's phenomenal horizons outwards from the concrete 'sapeurs' to 'two-thirds' to 'this country' and on to his wider construction, that is imaginatively overlaid onto the scene.

When considered in its totality, the MODE variable allows for a systematised account of CVR as a material and semiotic technology. I have used MODE here to describe the constraints inherent to the form and the choices made where the producers of the texts are negotiating the technical constraints of the form. This brings CVR's main conceit into focus in functional terms: that its indexical representations and self-erasure enable a virtualised context of situation for the spectator. Osambia-Kpwata Fyfy's moment of multimodal artifice in C#26 is evidence of this conceit in the expressed language choices used to facilitate the kind of situation Firth referred to where she 'prehends' the spectator's potential actions. In general, the texts exhibit a tendency to use more subtle means to foster a sense of dialogic, contemporaneous experience for the spectator. This kind of language dynamic will inform the text's other register variables, which I

will continue with a discussion of *field*.

8.2   Field

The texts were evaluated for their *field*; for what they are about. This includes their subject matter as well as their semiotic 'aboutness' as socio-semiotic activities. For example, Congo VR's field includes, but is not limited to its *reporting* on the *DRC's geography.* As register is in direct contact with genre, the discussion will be contextualised for how semiotic activities in the texts realise generic traits. Briefly, the texts are combinations of stories and Reports, with variation in kinds of subgenres. The discussion of field extends the discussion of MODE, as each text assumes different kinds of experiential distance, with fields that include the immediate virtual context of situation and more reflective constitutive kinds of experiential field. Both films are essentially didactic as they reduce the particulars of their field to an underlying truth, as the transformative effects of the LAP and the paradoxical nature of the DRC. The ways that the didactic voice manifests in the films is elaborated through field. As with MODE, however field is dynamic quantity and the kinds of subject matter and semiotic acts performed by specific participants will reflect the texts' 'distributing access to roles and activities' (Poynton, 1985, p.56). Field, therefore, provides cues as to the nature of the voices heard throughout the films.

8.2.1   Subject Matter in the Texts

The simpler task regarding field is in the finding the texts' 'aboutness', as their *second order field* or 'subject matter'. There are many mid-level discrete fields of activity that constitute the main topics of the films. These are pertinent to describing individual speakers who will have varying degrees of access to these subject matters. The subfields are derived from the relative contributions of all instances of participants in the texts, including *people*, *things*, and *places* (in the verbal and visual field) as well as *verbiage*, *mental phenomena*, and *noun-groups* (e.g., 'his empire of riches'). The lists below capture the variance in the films in terms of what the films are about. The lists are ordered according to proportional weighting of each subfield (numbered in parenthesis). For example, Mamie is the primary subfield in *Mamie's Dream* and what the film is 'most about' as she is seen most frequently and references herself repeatedly in the lexicogrammar. A full breakdown of the subfield and the taxonomy used can be found in Appendix F.2: Subject Matter: Congo VR and Appendix F.3: Subject Matter: Mamie's Dream.

Mamie's Dream	*Congo VR*
1. Mamie (33)	1. DRC geography (49)
2. Education (22)	2. The Production / personified spectator (33: when camera-drone is included)
3. Family (18)	3. Corruption (27)
4. Gender (16)	4. Historic conflict (23)
5. Personal conflict (13): *in the verbal field only*	5. Travel and exploration (20)
6. Social recreation and play (9): *in the visual field only*	6. Resources (17)
7. Authority figures (9): *in the visual field only*	7. Politics (17) *in the verbal field only*
8. Self-determination/actualization (7): *in the verbal field only*	8. Culture (17)
9. Celebration (5): *in the visual field only*	

Analysing this aspect of field in *Mamie's Dream*, showed that the LAP, as the underlying topic and motivation for the film's production is notable in its relative absence from field as a visual or verbalised participant. It does manifest in many of the other subfields (e.g., 'education'; 'self-actualization') but only at a higher level and with extra-textual knowledge of the film, which goes against the constitutive *participant-equals-field,* bottom-up process used. Also evident in Mamie's field is the visual-verbal divergence of subjects across the subfields. Of the nine sub-fields, three are represented in the visuals alone and two are presented solely in the verbal. The cross-modal divergence reflects the film's MODE as language takes a reflective role in her representations. In many of the shots, Mamie is constructing autobiographic experience in the voice-over while presenting complex re-contextualising relations with the visuals.

Congo differs where all but one of its subfields are represented both visually and verbally (politics and self-determination' are realised verbally, only). This reflects the many parallel structures, deixis and other semiotic resources used to position the spectator in an immediate cross-modal situational field. Also notable in Congo is the heavily weighted sub-field 'The Production/personified spectator' which highlights the reflexive nature of the text. Part of the film's field is the representation of its own textual production on different levels, as implicit where they are simply visible, and explicit when enacting the journey. Also significant in this regard is the spectator's overt incorporation into this subfield, explicitly in moments of direct address and personified perspectives, and implicitly by association in other shots (in other words, 'the team', 'I', and 'you').

8.2.2   Socio-Semiotic Field and Genre

The verbal circumstance and process types included in the texts' lexicogrammar provides necessary information on their specific kinds of semiotic field as well as how they realise genre. Together, they realise the texts' overarching generic goals, to bring the spectator into meaningful

contact with human experience while explaining the social worlds with which the films intersect. As might be expected of nonfiction texts, *reporting* is the dominant semiotic activity in both texts, but with variation in the kinds of reportage used. The texts also differ in their modes of address in the ways they *share* human experience and *expound* on the social topics. The texts are also dynamic in their use of field, where processes relate to *specific*, *general*, *actions* or *entities*, thus granting speakers differing kinds of access to field (cf. Martin and Rose, 2008). Taking the films in turn, I will discuss their semiotic activities for the way they realise genre, as *staged goal-oriented processes.*

The choice of verbalised circumstance in *Mamie's Dream* realises genre as the fusion of *storytelling* and *compositional Report,* as illustrated in Figure 29, below. The generic staging is realised almost entirely through circumstance of *time,[39]* which locates the spectator in a temporal flow of episodic events, construing matters historical, current, and future. In terms of staging, past time constitutes the *orientation*, *conflict*, *evaluation*, and *resolution* phases of her narrative, and present and future time signal the *coda* of her narrative, which is also the segue into her *Report* (see Figure 29: Mamie's Dream generic staging, below). The main story resolution (joining LAP) also functions as the establishment of the proposition of Mamie's Report, where she decomposes the LAP into its benefits, all circumstantiated in present and future tense. The Report takes the form of Martin and Rose's 'orbito-nuclear' structure (Martin, 1984) where a main entity is established (LAP) and subsequently decomposed (pride; recognition; respect; self-reliance; social influence; realisation of childhood dreams). The fact that the use of linguistic circumstance conveys so much of the text's genre points to the 'semiotic work' (Matthiessen, 2009) being done by the lexicogrammar and how the producers are using the virtual context of situation.

The text's choice of process types shows the kind of rhetorical strategy used to achieve the texts' generic goal. In her narrative, Mamie is predominantly reporting as *chronicling,* presenting a neutral account of events, to which she bore witness. By favouring action processes over relational ones, Mamie maintains a neutral reporting voice in her recounting of events. Her storytelling also exhibits a field comprised of *specific action*, as opposed to *general entities*, maintaining an experiential flow through her narrated events, that are simultaneously enabling the particularization of broader truths. Her narrating voice bears a didactive component which is subsumed into her personal voice in the first portion of the film with an exception in shot#3 where she introduces a circumstance of manner ('like many families in Sierra Leone…'), an inherently logical and enhancing circumstance that is noticeable in the way it breaks the tone

---

[39] *Mamie's Dream* uses fourteen instances of circumstance of time spread across eight shots.
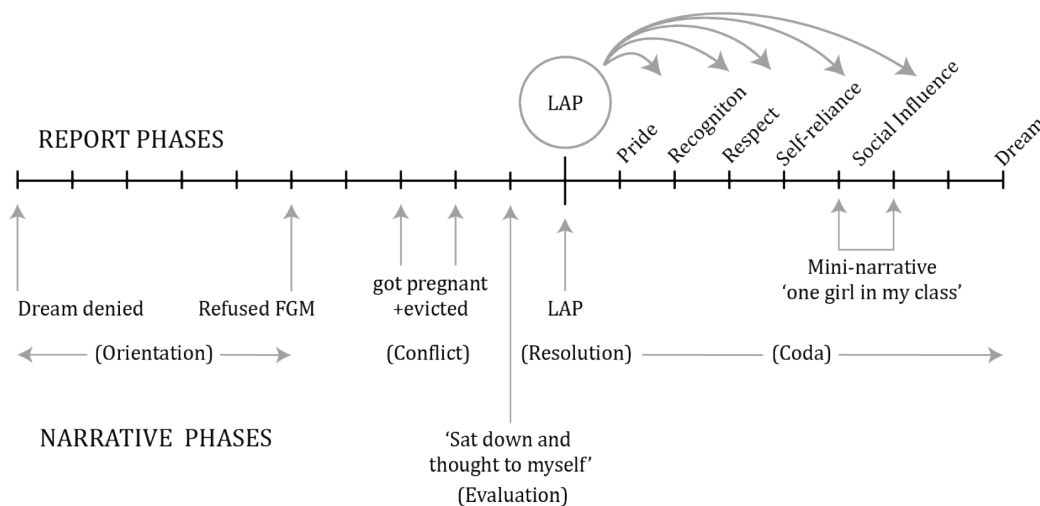
of the phase.

Figure 29: Mamie's Dream generic staging

Other than this, Mamie makes relatively few direct expounding statements in the first portion of the film, favouring second-hand reported commentary, such as her father's projected categorisation of girls' value, 'educating a girl is a waste of money'.[40] This keeps the spectator in the temporal flow of events, as her experience of hearing his opinions. Mamie does not make any direct appraisals of her fathers' opinions or actions, instead allowing his actions to speak for themselves. This approach also allows her to present relational processes, and general entities by rank-shifting them to verbiage or mental phenomena ('my father said, educating a girl is a…', 'I heard, it's painful'). In this way, expounding can be done without removing the spectator from the temporal flow of events, allowing her to make her commentary without risking her neutral voice.

When Mamie enters the Report phase, her semiotic switching from recounting to expounding, and from specific actions to general entities signal a change in her role in field, and the mode of address which becomes more didactic. Her perspective is replaced by a more objective 'higher' voice, as in 'a woman CAN be a teacher' and 'I am a community leader and a role model' which cast her as general entities compared to the specific actions in which she was previously bearing witness. Where her autobiography was feasibly in her own voice, a structured 'textual' voice is heard in this phase, where Mamie's voice is fused with the objectives of the LAP. This switching affects field by reconfiguring her place in field, from conduit in a reflective field, to component of a visible field that the spectator is privy to in their virtual situational context. Semiotic distance is contracted, if only by degrees where she now reporting on things that can be seen.

Mamie's Dream differs from *Congo VR* most significantly in its semiotic activity of *sharing*. By

---

[40] Reported speech is found in approx. 20% of all clauses, including direct and indirect quotes.

sharing personal values and experience with the spectator, Mamie's generic goal involves the spectator's access to the *meaning* of the events construed. Sharing is the second most frequent semiotic activity in *Mamie's Dream*, where Mamie is heard describing her personal reactions to and appraisals of things. A key moment of sharing in the text is in her self-realisation, 'I sat down and thought to myself, I am not going to give up on my dreams' which lends her interiority to the film deepening the narrative component. Sharing is used to allow the spectator to evaluate FGM and LAP through Mamie's eyes, beyond her 'reporter' role. This moment of sharing is matched only in a few other instances ('I was scared I might die', 'I felt so proud', 'I was alone'). Importantly, where Mamie is sharing *experience* as opposed to *values* she can present and evaluate her subjective experience without direct commentary. For example, 'I was scared I might die', is the sharing of a fact, as opposed to a hypothetical value sharing such as 'I was against FGM as it is a terrifying and unnecessary ordeal'.

Congo VR realises a different generic mix. It is similar to *Mamie's Dream* in the way it constitutes historic field but differs as a multivocal text that uses the immersed spectator as a text forming entity. *Congo VR*'s staging cannot be derived as easily from verbal circumstance, where place and time are used both to move the spectator experientially through the locations and to establish historic time. Alastair Leithead's prose also involves mixing and combining circumstance types, where for example he uses place a means of packaging valuable expositional information: "Travel is never easy in the DRC, especially *here* [place] in this sprawling mass *of more than ten million people* [quality]". *Congo VR*'s use of circumstance of *manner* differentiates the film in terms of its generic goals and as such, its mode of address. Alastair Leithead's didactic voice is heard in his explicit use of manner as both *comparison* and *quality,* such as in C#15:

> "The sapeurs display **the energy and brash confidence of the capital** [manner:
> comparison]. Two-thirds of this country are aged under 25. **Young and optimistic**
> [manner: quality], they're desperate to burst free of the country's troubled history

To appreciate how *Congo VR*'s field realises its generic mix, it is necessary to consider the text as comprising three overlapping strands, listed here, and illustrated in Figure 30, below:

1. *Report*: The DRC as encountered via the social actors' testimonies and surroundings
2. *Historic recount*: The public history of DRC, realised in Alastair Leithead's and Ilungama Ayanda's speech.
3. *Narrative*: The immersive journey experienced by the personified spectator

Strand one is a compositional Report with a main proposition decomposed through the social actors who each represent different aspects of the subject matter. The main proposition is that

the DRC is essentially 'beautiful but troubled', a contradiction of wealth, poverty, and beauty. This is decomposed through the subfields discussed above in section 8.2.1: Junior and Princess Mamicho realise *DRC geography*, *culture,* and *the production* (obliquely through the spectator's personalisation); Ilungama Ayanda realises *DRC geography*, *historic conflict*, *resources, and culture*; Osambia-Kpwata Fyfy realises *politics*, *corruption*. They differ slightly in the semiotic activities they use to convey their respective subject matter where each *report* on aspects of the DRC but have differing access to field and consequently a different kind of voice in the text. Princess Mamicho reports on general entities (Sapeurs) and their attributes; Osambia-Kpwata Fyfy also reports on entities, albeit more specified as *herself*, *the palace*, and *her relationship to Mobutu*. Neither of them reports on processes, which differentiates them from Ilungama Ayanda. He speaks in semi and complete generalities, eschewing his immediate field favouring instead the construal of historic activities (e.g., 'The white people, especially the Belgians *seriously mistreated them*') and entities (e.g., '*The suffering* started with our ancestors, and it continues into today'). Where his voice is a testimony to general historic events, the other two speakers are more objectified and emblematic of their subject matter.

The second *public history* strand realises a *historic recount*, as the constructed field is episodically staged. The colonial era is constituted entirely in Ilungama Ayanda's reporting, as well as Alastair Leithead's elaborations. Osambia-Kpwata Fyfy and Alastair Leithead use their language to construe the historic field *corruption* which is partly embedded in the spectator's virtual context. Osambia-Kpwata Fyfy references a context of situation in her *surveying*, as does Alastair Leithead. In Alastair Leithead's case, he also chronicles the unseen events (sipping champagne, beds going underground) in language which re-contextualise the visuals.

The third strand involving *the journey experienced by the spectator* is a kind of meta-field that incorporates the other two strands but adds generic characteristics of *narrative*. The film uses circumstance of *location* as *place* and *time*, to orient the spectator in the experiential flow of their immediate contextual field. This strand is notable in the way it mixes its genres: when the text is considered as a Report, this strand functions to decompose Alastair Leithead's underlying entity by moving the spectator to the various locations where the social actors 'await'. It also fuses subtle aspects of storytelling into the journey, including hints at conflict to be overcome (the rapids, difficult travel). These are admittedly trivial conflicts and not fully realised but are notable for their narrativization of the spectator in their immediate field. This strand is also reflected in the high proportion of *the production and personified spectator* subject matter that evidences the additional narrativized ways that the text incorporates the immersed spectator. The *guided tour* genre is also feasible in this strand, where in C#4 the spectator is initiated and oriented towards the coming journey, and where Osambia-Kpwata Fyfy and Alastair Leithead

use synthetic personalisation and deixis to direct the spectator's attention around the immediate situational context.
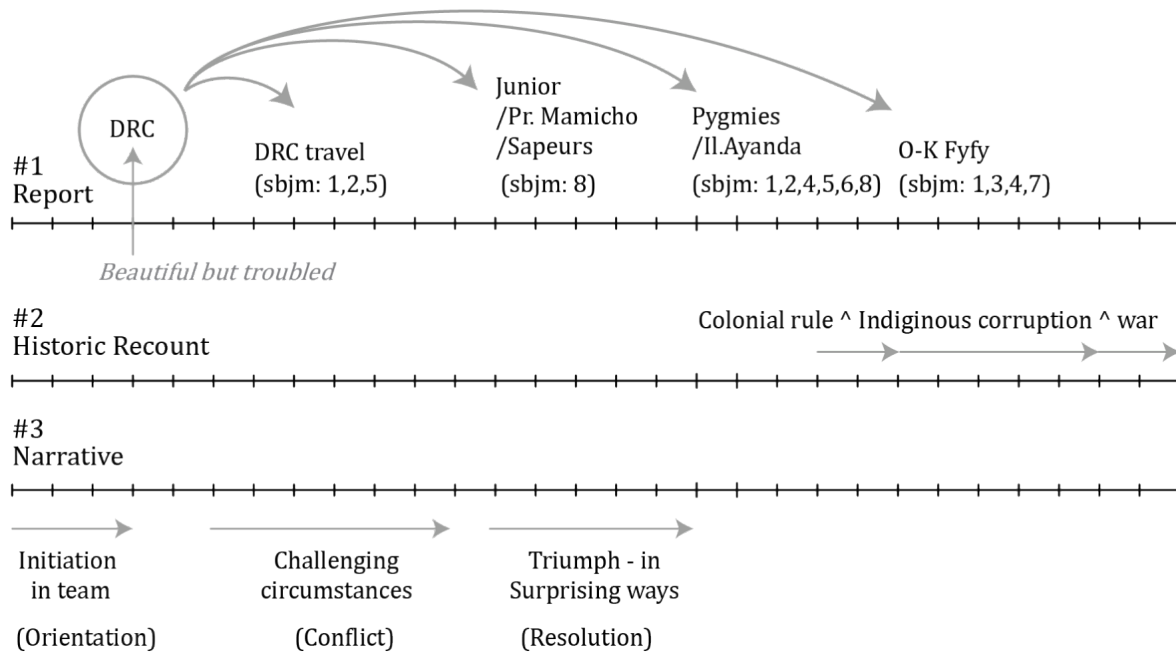

Figure 30: Congo VR's generic staging broken into three strands

The underlying didactic nature of the text is realised in *expounding* activities. Where *explanation as causation* is required in *Congo VR*, Alastair Leithead adopts a rhetorical device like Mamie, where he packages his commentary and exposition in objective reportage. He uses grammatical metaphor to do so, such as his swapping of agent and medium in the clause 'The railway finally opened up the Congo for exploitation' which makes the train itself the agent 'opening' the Congo, ellipsing the actual European agents involved, (e.g., 'the Belgians used the railway as a means of exploiting…'). In a similar fashion he transforms a relational process into a material one: 'the train to Kinshasa runs just once a week' which *categorises* the degraded status of the railway, as opposed to simply reporting on the actual frequency of the trains.

Regarding the spread of voices in the film, there is a hierarchy relating to the roles the speakers adopt and their access to the different kinds of field available. At the highest level, all social actors (and generic participants) constitute the subject matter 'DRC geography' to which Alastair Leithead has access to. He has the most access to field, where he can access the three genre strands and all of the subfields listed above. He has the 'voice of God' where his semiotic activities allow him to embed himself within and direct the spectator's immediate field (the journey). He is also an enhancing voice for the social actor's and their testimonies, where his socio-semiotic activities expound on complex metanarratives. He also has access to the field of historic experience, where he is allowed to co-construct historic grand narratives. In section 8.1.3

(Experiential Distance), I described Alastair Leithead's ability to use the immediate situation to expound on more semiotically distant entities. This reflects his playing with field also, where he is able to move freely between fields, embedding his language in the immediate field and transcending it to constitute a more distant field.

The social actors have a more restricted access to field and vary between speakers in terms of their roles therein. Technically, all diegetic speakers are part of the immediate situation that entails the spectator and their journey (strand 3), but it is their use of language that apportions their roles therein. Princes Mamicho's voice is embedded within the spectator's immediate contextual field, she both personalises and uses the semiotic activity of sharing to denote her access to the virtual field. Her voice switches abruptly to *reportage*, where she surveys her own attributes, categorising herself as a general entity and symbol of contemporary DRC culture. Her switching from subjective voice to objective voice is similar to Mamie in that it reconstrues her voice and her position in field. The switch is also reflective of a genre mixing that takes place in the text where the Report shades into a quasi-*tour*, in as much as it is not unlikely for a tour guide to interact casually with their clients before switching to 'officious' mode.

Osambia-Kpwata Fyfy differs in her access to field, as reflected in the way she uses MODE differently to the other speakers. In her direct address to the spectator in C#25, she is much like Princess Mamicho in her access to the spectator's immediate field, but without the overt reference to the spectator. She has access to a wider array of subject matter than Princess Mamicho, some of it is inadvertent however where she discusses 'Picasso's paintings', which are then re- construed in Alastair Leithead's wider field of *historic corruption*. In C#26, she exhibits a significant shift in her textual role with an access to field similar to Alastair Leithead's. By adopting a 'diegetic voice over' while directing the spectator's attention, she is embedded in the most immediate reflexive field overlapping with the 'production and personified spectator' available in the text. Like Alastair Leithead, she is thus allowed move freely between field.

8.3 Tenor

MODE and field described the texts indirectly for the way the spectator is constructed. *Reporting* and *sharing* for example presuppose a different kind of engagement with the spectator. Tenor realises genre explicitly in this regard: where the activity is *reportage*, it can be elaborated for the ways the text positions the spectator as a relatively passive and uninitiated entity. Where sharing is involved, the text draws the listener closer to make their own judgement on what is said and what it might mean for the speaker. This section discusses how 'goal-oriented processes' have interpersonal characteristics, through relations of *power*, *contact*, and *affective involvement*

(Poynton, 1985; 1991). As expositional texts, they share many underlying tenor traits, and will first be discussed as such for the 'global' tenor evidenced in both films. This will be followed with a more nuanced account of the texts, for how they differ and for how tenor manifests dynamically within each film.

The films' commonalities are realised in their MOOD and modality structures. Subject matter is *reported* in the texts through the indicative declarative MOOD form (with brief exceptions discussed below). This constrains tenor, constructing a one-way flow of information and a power relationship of *informer > informed.* In this regard contact is also maintained at a particular distance where the perception of imminent interaction is lessened, indicative of a relatively unfamiliar, infrequent relationship. As a consequence, affective involvement is also low as language-use evinces its scripted origins, negating any emotional component in the speaker's address.

Modality contributes to tenor in several ways. Visually, both films inhabit an indexical modality as photographic video, with a high naturalistic coding. Verbally, a general lack of modalizing gives the speech an overarching claim to objectivity, where speakers report without interjecting with subjective appraisals. The addition of music also determines the text's modality: where it is present as score, there is an implication that the producers wish to guide the spectator's emotional engagement excluding kinds of generic potential (e.g., straight news reportage). Also, both texts use voice-over, which can be construed as a particular kind of naturalism (or un-naturalism) when considered relative to the diegetic speakers whose voice and visage are fully integrated across the respective modes. When the baseline MOOD and modality are considered, the texts both construct a spectator with expectations of their role in the text, in other words, as receivers of information that will be augmented emotionally for them.

While modality is a global property of the texts, it is also modified in places. For example, *Congo VR* modifies the naturalistic image in its sporadic use of titles and graphic maps, and *Mamie's Dream* uses time-remapping effects to alter the temporality and augment the image. In both cases the author's hand is evident in the production, resulting in 'non-naturalistic coding orientations' (Kress and van Leeuwen, 1996, p.166). The converse is found in *Congo VR* which *naturalises* its scenes occasionally by omitting the voice-over (e.g., C#7 in the train station). Also, where musical score is omitted, the texts present seemingly unmodified experiences which are counterpointed with the more subjective multimodal representations involving the music (e.g., Mobutu's daughter and his ruined palace).

The texts differ most fundamentally in the way the spectator is personalised and visually

addressed. Mamie's spoken statements are complemented in the visuals, as *offered*: barring one ambiguous exception, neither Mamie nor any other participants acknowledge the spectator through direct verbal address or directed gaze. Lexically, her delivery is consistently neutral throughout with no interrogatives, breaks, continuatives or tagged clauses that would leave an imaginative space for the spectator to consider themselves as a notional interlocutor. Without an imagined reciprocity between speaker and spectator, *Mamie's Dream* positions the spectator akin to a 'close-bystander' looking in on her world. *Congo VR* is split evenly between visual offers and demands, where the spectator is addressed directly and included multimodally in the represented flow of events. Through deixis and other means of synthetic personalisation, the spectator is also given the perception of textual co-construction, and a structured role in the production. I will describe the construction of the spectator through their perception of power, contact, and affective involvement, taking each variable in turn and comparing the texts therein.

8.3.1   Power

Mamie's Dream modulates the power dynamic, in part by interspersing formal knowledge within more personal kinds of information. Her knowledge-status spikes occasionally, reminding the spectator of her expertise: "Like many families here in Sierra Leone, my parents wanted their girls to join the Bundu society and enter into womanhood". The construct is lexically dense and covertly assumes knowledge through a nominalized group "many families here in Sierra Leone' and verb groups 'join the Bundu society and enter into womanhood" (M#3). The first extract coincides with the first shot where Mamie is not visibly present, enhancing her objective lexical stance in Sierra Leonean matters.

Power in *Mamie's Dream* can be framed as evolving across the text where her upward trajectory of social status infers an inverse reduction of power in the spectator. Where the spectator begins as witness to her tragic back-story, they are subsequently 'equalised' as bearing witness to her triumphs. The fulcrum of this 'see-sawing' of power relations is found in M#9 in her evaluation: "I am not going to give up on my dreams - I am stronger than this". Prior to this, she is construed as relatively powerless, which manifests in several ways. First, her lexis construes second hand voices with perceived agency over her. Her father is an agentive force, getting to speak three times (quoted directly twice). With a voice in matters, he has textual agency which is copper fastened by his reported actions in M#7 ('threw me and my belongings out into the street'). Other voices are found throughout the rest of the text but are not quoted directly. Her father's voice is not heard in the second half of the film whereas Mamie is given additional voice (and bandwidth), in the introduction of her diegetic voice. In M#9 she is heard speaking a nursery rhyme to her daughter and in M#12 she is vocally interacting with the pupils. To augment this empowerment,

her diegetic voice is heard in a rare shift in MOOD, where she uses a modalized imperative, "*just* keep your faith […] Don't cry, *just* keep doing what is right").

The evolving power dynamic between Mamie and the spectator is also reflected visually, in the overall pattern of her physical activation realised in the kinetic mode, and in the relative spatial associations between her and the spectator. In shots M#1-9, Mamie is static, seated or both.[41] After Mamie's moment of realisation in shot 9, she is seen standing and active, with the exceptions of M#17 where she is seated but active through emphatic gestures, and M#19 where she is standing but static. The spectator's role can be typified as being someone expected to sympathise with Mamie in shots 1-8, to one who is subsequently made witness to her transformed status thereafter. This manifests in a novel 360° fashion in M#11 where the spectator is positioned in equidistant spatial configurations between Mamie and the teacher, placing Mamie on a par with him.

Congo VR is varied in the kinds of power dynamics between spectator and text. Explicit markers of power are found in Alastair Leithead's verbal authority in matters DRC. This authority manifests as both status and standing, as visually implied leader of 'the team and I' and guide for the spectator, as well as expert of DRC's geopolitics. Leithead's status as guide is realised across the modes in several ways. In C#4, the compositional structures in the boat 'point' to Alastair Leithead, marking him out as the principal figure in the shot as he verbalises his accreditation as 'BBC's Africa correspondent'. As a highly interpersonal shot, all modes combine to position Alastair Leithead as the leader, and the spectator as an initiate into the BBC team, by positioning the camera as part of the 'team' formation. This associative effect will be explicitly factored into later shots such as C#15 and 17, and with lingering effects in the other shots where the BBC team are not present. Structurally, his lexis in the boat in C#4 also realises power between him, the team, and the spectator through agentive actor-goal transitive relations where Alastair Leithead and crew are *actor* ('we'll be taking you'), and the spectator is *goal* (being 'taken').

AL's expertise is evidenced throughout the text in his formal lexis: when he is in exposition mode discussing complex matters, his lexis is dense allowing him to construct complex connections in a short timeframe. His speech in C#11 for example is very dense (63%) and his intonation comprises a high ratio of emphasised elements (SFL's *tone units*). The extract below is a single clause comprising six tone units indicating the multiple points of interest he is getting across in the clause.

---

[41] Static: M#1, 2, 4, 5, 8, 10, seated: M#1, 4, 7, 8, 9, 10, static and seated: M#1, 4, 8.

> "**Tra**vel // 1. is never /**ea**sy // in the /DR**C** // especially /**he**re // in this sprawling
> **mass** of // more than/ **ten** /million /people"

Through Alastair Leithead, the spectator is positioned as the beneficiary of Alastair Leithead's status. Any subsequent journeying and learning will be facilitated through him. His relative power in the text is also rooted in his access to the voice-over and its overarching textual construction where the social actors are heard predominantly in the diegetic soundtrack.

Junior and Princess Mamicho's power is attenuated to offer a different interpersonal construction. Junior's speech is brief and almost entirely interpersonal. His interjection 'hello!' sets a starkly different tenor to Alastair Leithead. which is maintained by Princess Mamicho. She articulates her standing on matters sapeur, while using multimodal resources to construct the spectator as one who is *on the level* with her. She verbally enumerates aspects of sapeur lifestyle, while presumably using her actions to point to herself as embodying *sapeur*. The camera-spectator maintains a vertical relationship throughout that is roughly at eye level suggesting parity of esteem. By tagging her lexis with a rhetorical interrogative ("A sapeur is an artist and in into fashion *you see what I mean"*) she subtly modulates her 'declarative power', implying instead that she prehends and is invested in the spectator's perception of her utterances. These resources combined construct the spectator as a ratified 'visitor' into the scene, suggesting she has granted them access into her world. Tenor is structured here in a larger sense to equate contemporary DRC as being on a par with the outside world.

Where Princess Mamicho's standing is inferred in her internal aspects, her visual status is less clear as others in the space seem ambivalent to her presence and actions. Ilungama Ayanda (the pygmy elder) on the other hand has unambiguous social status amongst to the other tribesmen in shot #21. Unlike Princess Mamicho however, the visual perspective through which the spectator encounters Ayanda is unbalanced as he is positioned below the spectator and thus visually subordinated. His visual status reflects the subject matter of his speech where he is chronicling his people's history of European subjugation. The pygmy elder's status is thus multifaceted: he has high social 'in group' standing but is weakened from the spectator's constructed 'outside' point of view.

Osambia-Kpwata Fyfy's power is attributed to her in a similar fashion to Alastair Leithead in her represented status as "traditional chief of Gbadolite" and her standing as Mobutu's daughter (presented in her graphic title). The relations enacted here are of interest on a more nuanced level, in the way her spatial configuration differs from the others. Where the other speakers' direct addresses were simulating 'conversational' interactivity, Osambia-Kpwata Fyfy's contact

is ambiguous. Spatially, her orientation and lack of visible address (by wearing sunglasses, she bars her gaze from view) suggest that she is to be observed as an object, as she performs her speech acts. Like Princess Mamicho, she tags her declarative but with a restatement of her own speech; 'It was a very beautiful *I told you, very, very beautiful'*. The underlying intent here is to signal her standing by reiterating the importance of *her* perspective on events, as something that only she knows, as witness to history. When taken as a whole, her distance and exclusionary demeanour combines contact and power to allow her to present herself to the spectator as an *object of power*.

Osambia-Kpwata Fyfy's differs from the other speakers in a structural sense regarding her use of channel. By allowing her to become a voice-over for the shot immediately following her diegetic speech act (C#26), she is allowed to describe the contents of a shot as *voice-of-God*. Her status is elevated beyond the others in her access to textual resources available only to her and Alastair Leithead. Further to this, she directs the spectator's perspective deictically through "in the entrance *to your left*", giving her a textual 'guide' status like Alastair Leithead's. Her ability to speak as voice-over and then to use that in a highly interpersonal manner places her high in the hierarchy of voice-types found in the film and positions the spectator accordingly.

8.3.2   Contact

Mamie's Dream enacts contact partly through the film's multimodal management of co-presence, as distance and socio-spatial associations. Visual closeness is used in *Mamie's Dream* to denote a kind of relationship, familiarity and thus mode of contact. As with the power variable, contact can be found to alternate across the two halves of the film, where the spectator is dynamically constructed as either *confidant* or *bystander* depending on Mamie's speech and actions. At the start of the film, the camera placement positions the spectator close to and alone with Mamie and contact, considered as a function of the expected relationships at such social distances, is at its highest (e.gs., in the boat (M#1); on the bridge (M#2); in the market (M#8)). Contact shifts around the film's pivotal moment in shot#9, whereafter the spectator becomes progressively distanced and shares the space more with others (in the associative parity relations discussed in section 6.3.1, *Relative Social distance and involvement*). This increase in distance instantiates a less intimate relationship, which is amplified by the spectator now sharing Mamie with other specified participants (her daughter and 'one girl in my class') who are given a degree of parity based on their associative distances to the spectator and Mamie. The overall effect on contact is to re- establish the spectator's relationship from one of *close-bystander* and *confidant* (while alone with her), to *one bystander among many.*

Ellipsis and substitution are resources used to establish contact, where the listener's prior understanding of context allows for certain things to be left out, and the assumed awareness of context assumes a degree of frequency in the relationship. Both texts adopt this device whereas *Mamie's Dream* is more nuanced in its approach. Mamie presents substitution in her lexicogrammar and visual ellipsis to set a particular expectation of the spectator. For example, in M#7 she does not specify the daughter's father in the lexis, instead presuming him as 'he'.[42] The effect is subtle but effective in imbuing an expectation of familiarity. Visually, the opening shot also assumes a relationship of sorts and a pre-existing knowledge of Mamie, for us to be included in such a casual and quotidian manner. Contrastive to Alastair Leithead's elaborate introduction in C#4, Mamie never identifies herself and ellipses all causes for our placement in the boat in M#1. The proxemics and use-space used to construct the interaction suggest that this is a 'typical' occurrence, without the need for any embellishment.

Congo VR works in a contrastive fashion, where contact is made with an explicit entity ('you'). More broadly, the spectator is constructed twice: structurally, through the full modal ensemble, and via synthetic personalisation, as a textual participant associated with Alastair Leithead and 'the team' in C#4. While the latter initiation into the team is explicit, the structural contact is more nuanced. Beginning with a high POV over a coastal area of DRC (with BBC identity overlaid), the film's opening sequence comprises shots that start at extreme heights and become progressively lower before 'landing' at water-level in the boat in C#4. The film begins with technical perspectives equivalent to the voice of the producers, enhanced by the *BBC NEWS* graphic and Alastair Leithead's voice-of God. The shot sequence functions to bring the spectator from global BBC perspective - into an embodied 'you' once landed in the boat. Familiarity is augmented by the generic synth music in the early shots that are comparatively non-African, compared to shots that show contemporary DRC life. The effect of contact here is of constructing a presumed and familiar 'outsider' POV, concretised and activated in C#4. This becomes a part of the narrative staging thereafter where Alastair Leithead, the BBC, and the spectator are cast as a collective of *outsiders*, making their way into the DRC. The effect culminates in C#6 with Alastair Leithead presenting the rapids (below) as a challenge they must face together, reinforcing the expected role of the spectator.

"For centuries, they blocked Europeans from exploring the heart of the Congo" (C#6)

Alastair Leithead's particular contribution to contact corresponds with his need to be both *expert* and *friendly guide*. Lexically, he remains in the indicative MOOD but varies the density of his

---

[42] As exophoric collocative reference: we see a young girl; Mamie 'got pregnant', we assume who 'he' is.

statements depending on which persona he wishes to convey. His formal lexis discussed above puts him at a distance to the spectator with a formality that gives way sporadically to more casual speech. In C#9 he opines, "this [train travel] is as luxurious as its gonna get", switching to an informal lexis ('gonna') that is indicative of a more casual and frequent relationship. This aside serves two complementary functions, to exhibit an insider point of view, of being 'in the know', while also appearing relatable to the spectator, constructing a common experience of letting the spectator in on his expertise. This coincides with a visually personified spectator constructed by the mise-en-scene (the seating area and water bottle).

Alastair Leithead draws the spectator close elsewhere through personalisation of both he and the spectator, through implied expectations of interactivity. His use of 'we' for example (C#4, 17) creates a sense of solidarity between the spectator and Alastair Leithead. His use of deictic reference both positions Alastair Leithead as textual guide, and draws the spectator close, as he and they are in an imagined common moment of textual production (e.g., 'The rich and famous sipped pink champagne by *this* pool'). In terms of contact, Alastair Leithead's switching modes always functions in part to enhance his status, as an effective guide to the DRC.

The social actors also present varying levels of contact, that correspond to their underlying textual roles. Contact is perhaps at its strongest for a fleeting moment in Junior's 'hello!'. Whereas other speakers personalize the spectator while informing us on some subject matter, Junior's utterance, as an interjection, is empty of subject matter and entirely interpersonal. He is highly interactive in both shots (C#13, 14) as is his mother. She remains at a very close distance (intimate-personal) throughout the shot maintaining a closeness that augments the parity of power established in the spectator's verticality, positioning them both as equals. The perceived roles are both close and, 'on each other's level', allowing for the signification of the cultural similarities of contemporary Congolese and their (visiting) European counterparts.

The nature of contact between the spectator and Ilungama Ayanda differs significantly to Junior and Princess Mamicho. As with all others, his speech is in the form of declaratives and the physical distance is roughly the same as with Princess Mamicho. His demeanour differs, however: his forceful gestures with the rubber-leaf create a vector and barrier between interlocutors, enhancing distance and suggesting a more antagonistic relationship. Additionally, he sits between the spectator and the other pygmy men placing the spectator in a spatial opposition to them, in a stand-off of sorts. As such, contact is minimized to the point of negation and the spatially isolated spectator's outsider status is enhanced. His speech and gestures take on a more pointed character implicating the spectator in the subject of his verbiage. Essentially, where Princess Mamicho endeavoured to bring the spectator close and into her world, Ilungama

Ayanda intends to maintain a discernible distance; keeping the spectator 'in their place' so to speak.

Osambia-Kpwata Fyfy also engenders a very low degree of familiarity compared to Alastair Leithead, junior and Princess Mamicho. I described her elsewhere in terms of social distance and involvement, for the unusual use of interpersonal resources and ambiguity in terms of her offer/demand status. This distance and lack of involvement is mitigated however in her speech as she personalises both herself and the spectator. Tenor becomes somewhat ambiguous, which is compounded in the following shot by the structural closeness of her deictic references and amplified synthetic personalising of the spectator. When she says, '*At the entrance on your left*, there were once Picasso's painting', there is a sense of textual closeness, where both speaker and spectator are constructing the text together.

8.3.3   *Affective Involvement*

Affective involvement (AI) assumes markers of affection between interlocutors, such as the tailoring of vocatives and terms of address (Eggins, 2004, p.102). This is of course difficult to apply to the texts, as they are of a very different genre to the conversational genres typically used to formulate AI. Another marker of AI is in the content of speech, which tends to be more contentious, the higher the affective involvement. We argue freely with those we are close to but maintain polite and general 'chit chat' with those we have low AI (ibid.). This is the more applicable dimensions of AI, where the two texts vary the kind of subject matter between speakers. It is also possible here to factor in the phonological aspect of speech, as intonation patterns open or close the texts to the potential for AI. As such, this aspect of tenor, as applied, gets close to the issue of empathy being felt by the spectator in CVR.

Mamie's Dream presents a pattern of self-personalisation and descriptions of highly sensitive matters which foster relatively high levels of AI in the lexis. Her intonation is at times at odds with the lexis however where she adopts a 'phonological paragraph' style of *paraphoning* (Tench, 2020) indicative of reading from a script. She does break from this on occasion however, presenting more natural moments of affect. Her enthusiastic, idiomatic "He didn't have to ask me twice" is amplified prosodically as a 'committed' pattern.[43] Her prosody also modalizes the otherwise neutral lexis lending Mamie a vulnerability not expressed in the lexis. In her sharing moment, 'I was scared[44] / I might die',[45] 'scared' is spoken with urgency and 'die' is softened to reflect just how 'scared' she was. Similarly, in M#8 her intonation varies similarly across 'My

---

[43] Mid to high pitch rise-fall.
[44] High-fall strong': powerful expulsion of air indicating forceful emotion.
[45] 'Low-fall mild': weak expulsion of air indicating mild emotion.

father <u>said</u> / I was a prostitute [...] I was <u>alone</u>". By using both pitch contours in succession, the spectator is privy to a range of emotional peaks and troughs in her speech. This combination of intonation patterns, presents an insight into Mamie's emotional world including a vulnerability in her both as remembered child and as contemporaneous speaker.

Congo VR differs regarding AI, particularly in Alastair Leithead's voice-over. First, the spectator is not given access to anything sensitive regarding his experience, and any contentious issues voiced are extrinsic, relating to the DRC. Also, he paraphones to a higher degree than Mamie, giving his delivery a polished flourish indicative of practiced oratory skills, rather than moments of affect. C#4 is an example where he begins his clause at a high pitch and slowly lowers it across the complex, before 'resetting' at a high pitch for the next clause complex. This is more typical of news reading than personal conversation (Tench, 1996). Even when he levels his intonation across the shots, his formal lexis remains a barrier to any affective involvement in the spectator.

It is difficult to use Princess Mamicho, or Ilungama Ayanda's speech patterns as they are speaking in their native language, translated to an English voice-over. The former participant's lexis construes attributes of the Sapeurs and is somewhat neutral in affect. The latter does speak of contentious issues, but it is difficult to reconcile this as enabling a high degree of AI, when the contact and power are considered. The perceived confrontational demeanour means that AI is ambiguous with regard to Ayanda. In a similar fashion, Osambia-Kpwata Fyfy's multimodal representation signifies a distance that is not conducive the high AI. Her lexis and intonation provide certain markers however that make involvement higher that Ayanda. For instance, her repetition of 'very beautiful' raises the exchange from the diatribe level seen with the pygmy elder, to a more spontaneous construction, that is admittedly at odds with her visual presentation.

When tenor is taken as a whole across the texts, it is found to be dynamic. In *Mamie's Dream*, tenor manifests as a sliding-scale that is concurrent with field. As her field becomes more generalised in her Reporting phase, power is increased, and contact is lessened. In *Congo VR* tenor varies by speaker but in a less consistent manner than *Mamie's Dream*. Alastair Leithead's tenor is predictable in the way power and contact oscillate with his situational needs and different roles in field (as Reporter and guide). The social actors' tenor varies largely depending on whether there is a requirement to verbally address the spectator. In this regard, Ilungama Ayanda is the odd one out as he does not explicitly acknowledge the spectator and as such is in a different bracket to the others regarding contact, where they personalise the experience. This in turn leads to the performative aspect of tenor. While it is useful to make analogies to actual relationships when analysing the speaker's tenor, it must be remembered that these are

synthetic and consciously performed to the camera. Visual tenor is enacted largely by the producers of the texts and the way they construct their mise-en-scene and spectator perspectives. These relate to Marquis' performative 'tiers' (2014) which I discussed in section 2.3.4, Nonfiction and Performance, and which I will address further in the chapter's conclusion.

8.4 Conclusion: Characterising Nonfiction CVR

Thus far, this chapter described the texts as the integration of the three systemic functional strata: as low-level *language* choices realising a *situational* context, which in turn realises the text's *genre*. Where the first research question is concerned, CVR was addressed as a semiotic technology, directly in its MODE and indirectly as the confluence of semiotic activities and performances reacting to and 'playing with' CVR's novel modality. This also addresses the sub-questions concerned with the configuration of discourse. I will summarize those findings here for their own relevance and use them as a frame for discussing CVR as a mode of nonfiction production, from a practical, ideological, and ethical point of view. To do so, I will re-introduce the concepts of performativity, semiotic autonomy, and semiotic control.

CVR's MODE reiterates at a higher level the observations made in the metafunctions relating to the aspect of simultaneity and rotation. At the level of language, meaning manifests as indeterminate and relative, as the real-time construal of experience and the structuring of the interpersonal spectator. This is contextualised in MODE as the novel mix of an asynchronous yet informationally loaded bandwidth, where the 360° image affords the spectator real-time access to information that would otherwise be segmented as an edit. The spectator's rotation and object permanence allow for reflexive contextualisation of speakers, paralinguistically as well as extralinguistically, including the spaces they inhabit. That the rotation is performed in the spectator's own reflexive and embodied time means that CVR's bandwidth affords a perception of semiotic autonomy, where contextualizing information can be subjectively and cumulatively selected in real-time, depending on the needs of the spectator to resolve meanings.

From a linguistic perspective, MODE also presents an insight into the way immersion and presence facilitate discourse in CVR. When the reality of audio-visual representation is bracketed, and the typically technologically oriented *virtual environment* is considered instead as *virtual situational context*, it is possible to elucidate the choices made by the producers of CVR texts. Producers script CVR with the assumption that the spectator will experience the 360° image as their own immediate surroundings, in which case speech acts are predicated on the use of language relative to the virtual situation. Alastair Leithead and Osambia-Kpwata Fyfy prove this to a degree when they 'play' with CVRs MODE in their highly contrived moments of deictic textual production. when Osambia-Kpwata Fyfy says, 'the entrance to your left', her

speech is scripted with the planning and foresight to imagine where the spectator might be oriented. This is reflected as the experiential distance between spectator and (simulated) context where the intention is to simulate a 'close' distance between spectator, speaker, and subject matter.

8.4.1   Semiotic autonomy

When register is considered in light of the lower-level semiotics discussed in the metafunctions, a tension emerges where the semiotic autonomy found at the lower level of language, becomes necessarily constrained at higher levels of context. This tension can say much about the nature of interacting with the social world in CVR. In earlier chapters, I described the spectator as having a degree of autonomy in the construction of low-level meanings in the experiential, logical and interpersonal functions, where the aspects of rotation and of simultaneity are concerned. The findings discussed in this chapter point to the overriding nature of what constitutes 'reality' at higher levels of meaning. To claim an objective experience of what is captured in the 360° image is to ignore the structuring effects described in the film's register variables where the spectator is socially constructed by the texts. On a fundamental level, to have an experience of a virtual context is to have access to an immediate *field* of experience and in texts such as these, that field is structured to realise genre, where the spectator's experience is based on the manipulations of the social world required to construct a history, a Report, or a story. Thus, any contact with 'reality' is heavily mediated by the needs of generic staging and the many semiotic activities required to maintain generic coherence. In the case of these texts, it is ultimately the overarching didactic voice that is heard in the actual voices, as the *organizational strategies* (Nichols, 1983) employed by the texts' producers.

8.4.2   Semiotic Independence

As the specific descriptions of the texts' MODE, field and tenor have shown, CVR experiences are constructed around a particular kind of contrivance, where CVR's unique bandwidth is manipulated to support a virtual, mediating context of situation. This in turn becomes the profilmic field of experience for those represented, bringing about a particular mode of performativity. The kinds of performances required to foster situational closeness, such as the use of dialogic medium and familiar tenor, point to the aspects of performance that are in response to the asynchronous limits of the form. A 'CVR performance', must add meaning to the spectator's immediate field and requires the speaker's *prehension* of the notional embodied spectator. This takes different forms depending on the channel that the speaker inhabits and the subject matter they need to construe. As voice-over, *Congo VR* performs to the embodied

spectator in an overt fashion, through deixis and synthetic personalisation with Osambia-Kpwata Fyfy taking it to a level not seen elsewhere in the texts. Mamie does so also in a more muted fashion: her shading from dense expounding to more dialogic sharing, and the resulting contraction of semiotic distance is a more subtle contrivance than Osambia-Kpwata Fyfy's but no less of a performance.

The performances required by the diegetic social actors also involves a kind of synthetic prehension of the spectator with the distinction that their own immediate profilmic field of experience is with the *CVR camera*, imagined to be the spectator. Textually, speakers such as Princess Mamicho are required not simply to perform their lines but to construct a field of experience that is largely centred on a non-existent entity. Where all filmmaking, as profilmic is essentially the overlapping of fields (the production of a film and the personal field of the social actor), the situation that the CVR actors are negotiating in their performance does not include an embodied filmmaker. In Marquis' performativity model, their semiotic independence is twice reduced. First, in the technical situation where they perform to the CVR apparatus with the further requirement to imaginatively prehend a spectator. Secondly in the overarching generic staging of the texts where their testimonies become subsumed into the organising voice of the filmmakers.

Alastair Leithead's performance differs, reflecting his access to, and manipulation of the films register (such as casual, formal, dialogic). His voice of God status is not simply a function of his ubiquitous voice-over. He has access to all fields represented in the text, which involves the embedding of his own voice: addressing the spectator in their contemporaneous field, chronicling the pygmies hunt through voice-over, and weaving an experientially distant socio-historic narrative. Also, Alastair Leithead's voice is not restricted to one kind of field at a time, rather he 'register-switches' between reporting on the here-and-now while enhancing it with commentary. He is thus performing multiple parts but his significant effect on the text as a whole is in the constraining effect he has on the social actors. This is found to have an ethical dimension when Osambia-Kpwata Fyfy's descriptions of the former majesty of Mobutu's palace (her field) are recontextualised by Alastair Leithead into his 'corruption' subject matter (his field).

Through register it is also possible to articulate some of the constraints placed on the producers of reflexive, participatory texts such as *Congo VR*. As I have discussed, the 360° image facilitates the spectator in comparing entities across the shot and these entities can be found to enact different kinds of field that are not always congruent. For example, C#15 incorporates two fields realised by two episodes: the sapeurs and the BBC crew, the latter including the spectator by

association. The sapeurs enact their subject matter as part of the overarching Report, as emblems of *modern DRC culture*. Simultaneously, the BBC crew are enacting their own reflexive subject matter, *the BBC production*. A schism presents where the sapeurs, aided by Alastair Leithead's voice-over perform their parts *on-stage* literally and also in the Goffmanian sense, while the BBC crew are enacting a relatively *backstage* performance. On one level, the scene indexes a social event, as a faithful simulacrum of the overall profilmic situation, where the sapeurs perform to the crowd, BBC, and camera. On a textual level however, it is a complicated mixture of modalities where the seemingly higher degree of naturalness evinced by the BBC (chatting, drinking) conflicts with the sapeurs textual performance. The converse is found in C#8 where Alastair Leithead is counterpointed by an episode of train passengers in their moment of natural interaction, who serve to highlight the otherwise artificial and performed nature of the shot. This is a high-level effect of CVR's simultaneity, where contrastive modalities coexist in the 360° image, highlighting each other by their differences.

9. Conclusion

This chapter is structured to address the research findings concerning 1) CVR, as a semiotic technology that configures discourse in a unique manner 2) the ramifications for the producers of nonfiction CVR texts, and 3) the methodological and analytical lessons gleaned in the process of the doing the study, that suggest new ways of approaching CVR analysis.

Taking these in turn, I will first discuss CVR for the ways low-level discourse is constructed in the intersection of the CVR image and the spectator's rotational engagement. This is framed by the higher-level semiotics of the form, first where CVR's simulacrum creates a virtual situational context and site of discourse, and secondly for the mediating effects of genre production on empirical reality. CVR is then discussed from a theoretical standpoint, adding to the debates surrounding CVR's claims to transparent immediacy.

Following this I will discuss the findings of this study by framing CVR as a mode of nonfiction production. This expands on prior discussions of the spectator's *semiotic autonomy*, the author's *semiotic control*, as well the filmed subject's *semiotic independence*. The discussion will incorporate theoretical as well as more practical implications for the CVR nonfiction producer.

I will conclude the chapter by addressing the study in terms of the systemic functional methods that I adapted for use in CVR analysis. This includes practical, process related aspects of the study as well as theoretical constructs required to conceptualise CVR from a systemic

functional point of view.

9.1 Findings

Broadly speaking, CVR's transparent immediacy can be described as a semantic phenomenon, that is largely owing to the way the rotating spectator constructs low-level contextualising meanings as they rotate around the 360° image. 'Contact' with reality is the sense of inferring meanings reflexively in real time, on a pre-recorded audio-visual artefact. Presence and co-presence in the virtual space are also semantic process of inferring, through rotation, the interpersonal semantics of the environment. The spectator has a degree of semiotic autonomy that is real but illusory in its scope. When CVR nonfiction texts such as the ones studied in this thesis are considered at the higher levels of discourse, they reveal a heavily mediated version of reality that is predicated on specific modes of performance. To enact a field of experience for the spectator, the social actors must partly forgo their own, leading to highly synthetic situations and a lessening of their semiotic independence. I described this in the previous chapter (section 8.4.2, *Semiotic Independence*) in the context of the diegetic speakers who are removed from the presence of an embodied filmmaker and required instead to instate a fictional interlocutor in place of the CVR camera. Similarly, nonfiction producers script their language acts to uphold a virtual context of situation for the spectator, as was found in Alastair Leithead and Osambia-Kpwata Fyfy's use of deixis in the voice-over to close the semiotic gap between speaker and listener, all imagined to be in the same space. Essentially, when CVR is considered across the levels of language, context, and genre, it becomes apparent that the immediate autonomous experiences with reality in CVR, while semantically 'real', nonetheless serve as a smokescreen in the wider authorial construction. To borrow from Waugh (2011), this appears to be CVR's ideological 'camouflage' as a transparent form.

Taking the research questions in turn, I will begin with the following question and its sub questions:

1. *What kind of semiotic technology is CVR?*
    a. *How does the CVR camera configure discourse in a manner different to framed imagery?*
    b. *How does CVR discourse unfold across CVR's full modal ensemble?*

At its core, CVR is seen as a contextualising technology that compensates for its synchronous limitations by presenting an information rich image that is pregnant with meaning potential, and where specific meanings are constructed by the spectator's sensorimotor sense-making

capacities. I described these meanings on a metafunctional level across various sections in chapters 5, 6, and 7 as representations, as interpersonal relationships and as structures of information.

As a representational form, CVR presents an omnidirectional, simultaneous field of activity, with a multiplicity of construable meanings. Discourse is configured in real-time where meanings are, like most audio-visual texts polysemic, but also open to being reflexively augmented, modified, and logically inferred. As I discussed in section 5.1.1 (Transactional Processes), from a transitive perspective, participants were found to be *converted* dynamically from actors to goals to mental phenomena, each with subtle implications for their status in the overall higher-level activities. Mamie, for example was frequently transformed by her surroundings, often placing the underlying message at risk (described in section 5.1.1). This multiplicity reflects the 360° image as being inherently logical and enhancing, where the rotating spectator compares and connects disparate entities, necessarily circumstantiating and transforming their relative meanings.

From an interpersonal perspective, CVR's omnidirectional space organises human figures and episodes, connecting and disconnecting them from each other, and the spectator. The camera, as proxy, necessarily incorporates the spectator into relations with multiple entities simultaneously, requiring them to negotiate their social position in the scene through reflexive comparisons with others in the shot. This presents a significant disjuncture with conventional framed filming, as the typical interpersonal systems of social distance, involvement and perspective are available in a conventional sense, but critically they are also subsumed into systems of *association*. Where interpersonal relations were hitherto enacted bidirectionally through the image plane, the 360° image requires an awareness that to be with one person or group is to be necessarily distanced from another. This has a potentially mitigating effect on filmic identification, where *relative solidarity* becomes a foregrounded feature of the form.

Multimodal meanings become novel from a CVR perspective in the establishment and maintenance of the virtual immediate field of activity experienced by the spectator. This was described as part of CVR's materiality, where its transparency is partly the grafting of filmed situations onto its users' field of view conflating both as a novel virtual situation of discourse and experiential field. I have illustrated *Congo VR* as a reflexive text, comprising a raft of linguistic devices to maintain a semiotic distance indicative of a real situation that interlocutors would find themselves in in physical encounters.

The rotational aspect of CVR's textual construction suggests that the meanings inferred in CVR

are inferred in the spectator's reflexive sense of time, as their own phenomenal horizons. With this comes a particular kind of *semiotic autonomy*. It is this autonomy that characterises CVR as a transparent form and is the semantic aspect of the claims of *direct contact* with reality, where contact is not with the filmed subject but rather in the process of inferring contextual meanings from the situation in real time. The ideological implications of this can be teased out further via the second research question:

2. *What are the theoretical and practical implications of question one for producers of nonfiction texts?*
    a. *How does the nonfiction CVR text construct the nonfiction spectator?*
    b. *How can CVR nonfiction texts represent the social world?*
    c. *What are CVR's affordances and limitations for CVR documentarians?*

From a theoretical perspective, the study adds clarity to the experiential aspect of CVR, which allows for the constructed CVR spectator to be considered on an ideological level. The spectator's semiotic autonomy hints at how CVR is open to reiterating the ideological tropes of other transparent nonfiction forms, namely that the viewer can have direct and meaningful encounters with the filmed subjects. 'Meaningful' as elucidated in this study is semantic, as the spectator's reflexive decoding of the scene by rotating their view. Importantly, any semiotic autonomy that CVR affords is in the relatively low levels of meaning-making in the diegetic 360° space, as captured in the metafunctions. When these meanings are considered in the context of the social production of text, any 'reality' is essentially a *textual* reality, determined and constrained by the situation of filming and by the necessities of genre. On the most fundamental level therefore, the spectator is constructed with a sense of agency that is exaggerated in scope. To be 'greeted' by Princess Mamicho is to be in contact with a highly contrived profilmic performance that has the trappings of 'reality' - in the ad hoc and natural activities that surround her, which the spectator 'discovers' as they rotate their view.

Analysis of the texts' register illuminated the way the spectator's low-level autonomy is constrained by semiotic activities that are both in the service of maintaining the virtual context of situation and in realising nonfiction's generic mix of storytelling, narrating, reporting, expounding, and in the case of Congo VR, (tour)guiding. Both aspects of register require manipulations of the social field, shaping the reality available to the spectator. For example, reality was manipulated in the texts as storytelling to fulfil the generic conventions of conflict, evaluation, resolution and coda, and as Report to organise reality into a series of particularizations of deeper truths. What is particular to CVR is the way both factors overlapped,

where the immediate virtual situation enabled by the texts' MODE was at times incorporated into the generic construction of the texts. *Congo VR* in particular used reflexive and participatory tropes to mesh the immediate virtual field with the broader exposition, to the point where the text shaded into novel genres akin to a guided tour. In this regard, the reality claimed as transparent is itself the performance of genre.

Regarding the representation of the social world in the texts, the filmed subject was considered for the impact of CVR's production on their semiotic independence, and the degrees to which the texts imposed 'CVR performances' on them. When considered through the dramaturgical frames discussed in chapter three (section 2.3.3, *Nonfiction Frames For CVR*), performance was imposed across Marquis' three performative tiers (2014). Both films suggested that the notional spectator was *prehended* as an entity in their immediate field of activity, albeit in different ways. *Congo VR*, as a reflexive text required its social actors to synthesise their own field with that of a notional 'visiting' person. This contrivance reduces their 'everyday' performance through the impact of the camera (ibid.). In this regard, it is a predominantly technical performance that is entirely in the reproduction of a field of experience that prehends the existence of a notional entity, as heard in Osambia-Kpwata Fyfy's "in the entrance *to your left*". What is missing in the CVR performance is the embodied filmmaker who would otherwise serve as an interpersonal entity in their profilmic field of activity, and a mediating body through which the viewer could be subsumed. *Mamie*, as a non-reflexive film was more restrained but no less performative. Where *Congo VR's* subjects are *performing to* the camera, *Mamie* is *performing for* the camera (Waugh, 2011). Her performance was no less constrained by the camera, if perhaps in a more conventional fashion.

The affordances and limitations of CVR for nonfiction producers can be derived largely from the observations made in the metafunctions. CVR's simultaneity and multiplicity puts meanings at risk in ways unlike in framed filming. The aspect of semiotic autonomy for the spectator can be considered also as a matter of semiotic control for the producer. The relative indeterminacy of low-level meanings that manifest in CVR's simultaneity raises the possibilities of unintended consequences. I catalogued a number of these in the metafunctions involving a kind of 'long-tail' semantic effect across space and time in the 360° image, where entities contextualise each other leaving vestiges of circumstantial meanings as the shots unfold. This is an affordance in the sense that it replicates our experience of unmediated perception but becomes a constraint when the scene is not managed efficiently, leading to the risk of readings contrary to the preferred ones.

The interpersonal stakes are somewhat higher, where the associations made by the spectator have an ethical dimension, and where *in-groups* and *out-groups* are liable to manifest. Shades of

this were found in Congo VR where the personalised spectator was structured at times as a component of the BBC team, with ramifications for the kind of identification facilitated with others, such as the sapeurs and the pygmies. This point can be argued from both sides, as it presupposes conventional filmic identification. *Congo VR,* by choosing the reflexive mode is unabashedly making the spectator a personified member of the 'team' in-group, and therefore acknowledging the limits of identification with others in the contrived situation of filming.

Where the camera is considered for its placement as a physical apparatus, there is an apparent conflict in terms of what the producers need to do and show, and what the semantic repercussions are for their decisions. For example, there were examples in both texts where the camera was placed central in the mise-en-scene, presumably to afford the spectator equal access to details across all filmed subjects. Cinematically however, there are no semantically 'neutral' positions in which to place a reflexive CVR spectator, and to try to do so poses semiotic risk. A centralised camera becomes meaningful metafunctionally as the spectator must rotate their view to connect entities, resulting in logical comparisons between split-salient entities, which in turn present multiple branching exit points in the shot's reading path. Connected with this is the issue of perspective and contact when filmmakers ignore the semantic repercussions of the camera placement. 'Objective' perspectives are significantly mitigated in CVR, where the spectator is negotiating a flux of relative subjectivities in CVR imagery. Accordingly, to place a CVR camera for practical over semantic reasons risks it intersecting with action vectors and transactional segments, resulting in the spectator taking a transgressive perspective incongruent with the overall intentions of the shot. It is necessary therefore to foreground the socio-spatial dimension of filming, and not presume that the sense of presence in itself will translate into an ideal spectator who is necessarily receptive to the film's proposed meanings.

The third question involves the analytical process itself and asked the question:

3. What augmentations to multimodal methods are required to analyse CVR texts?
   a. What additional analytical frames are required to bring current multimodal studies into immersive domains?

Getting to grips with CVR's semantics and its configuring effect on discourse reflexively informed several adaptations and reformulations to SFMDA's systems. Figure 31, below on p. 209 illustrates the systemic aspects of the analyses where significant adaptations were made, or where re-evaluations were required in terms of multimodal meaning-making. These are presented from a functional perspective, but an observation made throughout the analysis was that CVR presents fuzzy boundaries between functional meanings: the textual becomes

interpersonal by the spectator's centrality in the structuring of information; the experiential function is pervaded at all times by the logical, and the ever-present visual circumstance (setting) contributes necessarily to all meanings.

On a fundamental 'grammatical' level, analysis was aided by reformulating of the rank system, to account for the simultaneity presented in the image, as well as the rotational aspect. This was partly the reaffirming of the usefulness of other rank systems developed by O'Toole (1994) and O'Halloran (2004) with modifications to account for CVR's field-of-view viewing mechanism. Other kinds of textual units were redeployed in CVR such as the conjunctively relatable unit (CRU) (van Leeuwen, 1991) that was useful as a generic unit of comparison.

More specific adaptations were required at the metafunctional level, many of which were somewhat common-sense, as they simply extended extant SFMDA resources without any requirement to re-evaluate their fundamental functioning in CVR. Framing, for example translated Kress and van Leeuwen's original conceptions of information arrangement from the two-dimensional picture plane to the 360° 'sphere', with an added interpersonal focus involving the spectator but performing the same essential compositional function. Other systems became more destabilised in the simultaneous space such as visual information value systems, which were open to potentially exponential and reductive meanings. Some systems bristled with their application in CVR where CVR's essential modality was arguably incongruent. Reading paths were a case in point, where the system required a linearity that in practice could be supported by CVR's many vector-based visual resources but presupposed a kind of engagement that is at odds with the more logicalizing nature of CVR. Connected with this were systems of salience, which also supported and enabled kinds of rotations and paths, but which were not always commensurate with the systematizing of the reading path. More broadly however, as a multisystemic approach the study can augment CVR works in attentional engagement, such as Brillhart's (2018a) and Brown et al. (2016a) by adding the dimension of cohesive *meaning* to attentional vectors.

Perhaps the largest proposed addition to SFMDA analysis with unique benefits to CVR, was Adam Kendon's socio-spatial models. These were incorporated into the study to manage the interpersonal complexity of the image and as a way of quantifying the associations made by the spectator. The models were effective in adding a layer of description to many of CVR's interpersonal resources, such as proxemics, perspective, involvement, and social distance. *Use-space* became effective short-hand for describing the confluence of transitivity, interpersonal relations, and the organisation of space, and as such transcended the interpersonal metafunction.
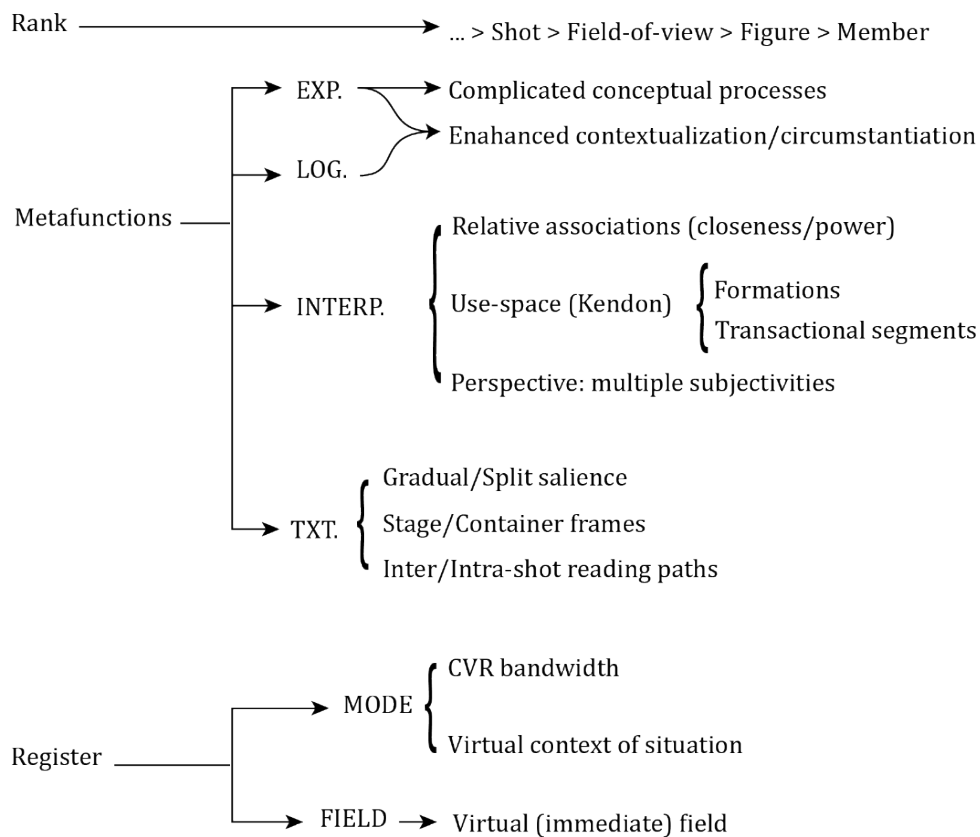
Rank ──────────────────────────────→ ... > Shot > Field-of-view > Figure > Member

Metafunctions ───┬──→ EXP. ──┐──→ Complicated conceptual processes
                 │   └──→ Enahanced contextualization/circumstantiation
                 ├──→ LOG. ──┘
                 │
                 │   ⎧ Relative associations (closeness/power)
                 │   ⎪
                 ├──→ INTERP. ⎨ Use-space (Kendon) ⎰ Formations
                 │   ⎪   ⎱ Transactional segments
                 │   ⎩ Perspective: multiple subjectivities
                 │
                 │   ⎧ Gradual/Split salience
                 └──→ TXT. ⎨ Stage/Container frames
                             ⎩ Inter/Intra-shot reading paths

Register ───┬──→ MODE ⎰ CVR bandwidth
            │   ⎱ Virtual context of situation
            │
            └──→ FIELD ──→ Virtual (immediate) field

Figure 31: Principal augmentations and additions to SFMDA systems

9.2 Contribution To The field

This study considered CVR experiences as semantic encounter with text. By taking a social semiotic approach, I have added descriptive clarity to CVR as a mode of nonfiction practice whose discourse is configured by a novel technology. I set out to clarify CVR as a semiotic technology and by using SFMDA principles I have tentatively conceptualised the technology in terms of its production and reception. I have presented an alternative semantic means of considering the spectator's rotating engagement in the VR headset. Fundamentally, from a nonfiction perspective, I have provided an alternative way of considering CVR which is often presented as a technologically determined experience of 'reality'. In doing so I have posited several adaptations that assist in extending SFMDA into relatively uncharted immersive territory.

At the time of writing, it appears that there are no systemic functional studies that interrogate the form at this level. I have referenced many studies being carried out that address CVR in a manner complementary to this study, each focusing on a particular aspect of the form. I have attempted

to subsume them into a comprehensive systemic account of CVR. 'Directing attention' for example is reflected in salience, reading paths and cohesion. Other CVR studies were formative as wider conceptual frames, such as the idea of media witnessing, and the potential for propagandising in CVR, as theoretical currents running through much of the study's findings and discussion, where the spectator was foregrounded for their meaning making potential in CVR.

At the highest level, the study successfully integrated social semiotic theories, whereby texts can be evaluated as a means of establishing social, contextual meanings. The systemic functional principles used to operationalise the social semiotic methods proved that SFMDA is indeed flexible enough as a meta-theory to cater for evolving media. In terms of the analysis portion of the study, the study drew from diverse theoretical SFMDA strands: SFL provided a stable grounding for analysing speech acts; visual grammar and multimodal theories of film and action proved apt as the basis for formulating analysis of CVR's novel spatiotemporal mix. The discussion phases of the study drew from the dramaturgical principles, originating with Goffman's original micro-sociological thesis and brought into the nonfiction domain by Marquis. The study successfully integrated this theoretical framework, which proved useful in conceptualising CVR's textual nature.

I consider the study as benefitting two groups, namely the producers, and the analysts of CVR. Where the producer is concerned, I have presented a broad theoretical consideration of CVR, as context for its claims to transparent engagement with the world. These kinds of considerations are important for the lower-level decisions made when conceptualising their productions. In a broader sense, by adding to the debate about CVR, this study could inform the overall developmental trajectory of the form. I have also provided grounded and practical considerations concerning the CVR camera and its use in social representations.

Regarding analysts of CVR, I have at the very least provided material for discussion of how immersive forms like CVR can be described from a systemic functional perspective. Many of the adaptations made to the extant SFMDA systems were practical and from my perspective were the product of this being the first foray into the field.

9.3   Limitations of The Study

The limiting factors in the study include the narrowness of what was conceived as 'nonfiction', reflected in the small sample size used to study the form. Owing to such limitations in scope, it was necessary to take a somewhat simplistic view of nonfiction - both conceptually and as genre. I am fully aware that these films, while somewhat typical of CVR's nonfiction output to date, are

representative of only a particular kind of text. When the broader canon of texts is considered, there will be exceptions to the kinds of observations that I have made within this study. Indeed, there are nonfiction CVR texts that look very different to the kinds that I have included, that most certainly do not subscribe to the kinds of nonfiction tropes that I critique. As such, I acknowledge that this study remains useful at a preliminary and basic level, towards teasing out and ascertaining the semiotics of the form. Throughout the study, I have aimed to keep the tone reflective of this limitation, focusing on genre primarily from a linguistic sense and avoiding making totalizing conclusions regarding nonfiction.

From a systemic functional perspective, the novelty of the study presented some fundamental challenges. Lacking a singular set of systemic resources to draw from, I occasionally required a return to the 'root' SFL system to illustrate and make sense of the findings. As such, I found myself at times lexicalising the 360° image and its effects; a phenomenon which is anathema to multimodal studies. This was found in the experiential chapter mostly (chapter 5), where linguistic phenomena (e.g., 'down-ranking') best described the effects of circumstantiation.

In terms of the study's modal units, the initial intention was to approach the texts from a more 'base' modal level. The specific modes transcribed and coded in the study were selected for their inherent spatial qualities and for what they could specifically say about CVR. They were ideal as a means of framing the data but as the study progressed, analysis increasingly devolved to discussing 'image' and 'voice-over' and at the higher levels, namely 'language' and 'field'. This was the natural revelatory nature of the study, as the SF model elucidated the form. Connected with this was the necessity to leave out detailed analysis of music as a contributing mode of engagement. This was purely a matter of scope, where the addition of music would have had a multiplicative effect on analysis. Also, music carries its own contested semiotics, and would require significant contextualizing throughout the study.

Other limitations were couched within the tentative nature of many of the observations. For example, Kendon's use-space was described for the impact that it has on the spectator's sense of perspective and where they transgress use-space, that it has a disruptive effect on 'reading' the shot, interpersonally. These observations are gleaned from general film theory and presuppose a specific kind of filmic engagement with texts. This is perhaps the nature of studies such as this, where it is necessary to evaluate a novel form in terms of tropes that are inherited from older forms.

9.4   Future Research

This study suggests several lines of inquiry suited to experimental research. For example, the use of Kendon's schema as an organising framework in CVR can be tested in the field to see how the phenomena I have posited, hold up. Similarly, the effects of cohesion and logical inference can be studied through user feedback. The study has presented CVR as a logical textual form, where engagement is predicated on sense-making. I would consider the studies currently looking at low- level sense-making paradigms in VR (e.g., Passmore et al., 2017) to be naturally conducive to studying CVR from a textual perspective.

Many of the frames and practices used in this study could also apply to the study of other extended reality (XR) forms, such as augmented reality. Such forms share the common thread of *embodiment*; requiring and utilizing rotation and immersion in the structuring of their discourses. As such, much of the linguistic framework used here, as well as Kendon's schema, could be brought to bear on a wider array of immersive technologies.

SFL and SFMDA are commonly used in critical discourse analyses. I feel that when the systemic adaptations I have presented here would provide a grounding for similar critical analyses of CVR texts, while also providing an opportunity to widen the study into other nonfiction uses of CVR. This study would provide more semantic grounding for critical studies relating to such things as humanitarian representations. Similarly, the study provided a glimpse into the uses of CVR in corporate and institutional news. I would consider it worth examining the ways that specific entities, such as the BBC adapt their practices and output from institutionalised conventional formats into more novel forms, such as CVR.

9.5   Closing Summary

The study presented a picture of CVR that adds descriptive clarity to the technology, and the ways that discourse is constructed between author, spectator, and filmed subject. CVR's uniqueness lies in the camera's ability to index situational contexts and present them in the headset as virtual fields of experience. The spectator is presented with an array of simultaneously occurring phenomena that is reflexively interpreted in real time. CVR is in this regard, a logical and enhancing form where the spectator is actively involved. It is the sensorimotor 'sense-making' aspect that defines the form, with real implications for the construction of text. When taken in context however, the reality that the spectator engages with is heavily mediated by profilmic requirements, and ultimately by genre. The technical affordances and low-level semiotic autonomy become mitigated as the text is framed at higher levels of social production. This

highlights CVR's unique formulation of older ideological claims to immediacy that falter under scrutiny. CVR is no less reliant on textual performance and if anything requires more technical and stage-managed performances than framed filming. With this comes a particular reduction of semiotic independence in the filmed subject as they must negotiate a spectatorial 'ghost' in the moment of production. Understanding the ideological claims made of CVR and the realities of textual construction will benefit the producers and commissioners of nonfiction CVR texts leading to the suitable and sensitive treatment of social subjects.

Through an in-depth analysis of two CVR texts, I have provided a semantic grounding for these claims by probing CVR discourse from the low to high levels, showing the interfacing of the minutiae of real-time construction of meaning with the social and ideological contexts of production. In doing so I have both characterised CVR as a semiotic technology and contributed back into the canon of systemic functional multimodal theories and practices. This has practical benefits for nonfiction production. It is not unlikely that CVR will morph many times in the future, as it finds its 'second birth' as a stable form. On a practical level, knowing the low-level semantics of the form will also inform the methods of production leading to more cohesive experiences.

This study also tentatively presents a system for studying CVR and lays the grounds for further studies. This involves testing the strength of the study's findings through other means and by extending the methods used here into other nonfiction CVR texts. Beyond this, the methods elaborated in this study should have relevance in other immersive contexts. Nonfiction producers will continue to experiment with novel immersive forms and this study will contribute to understanding those forms and the choices made in their production.

10.   Bibliography

360° VR Skydive for Next Level Indoor Skydiving. 2018. Perspektive Media GmbH Available at: <https://www.youtube.com/watch?v=n-55ww4JsuY> [Accessed 23 April 2022].

Aijmer, K., 2016. Modality and mood in functional linguistic approaches. pp.495–513.

Albright, M., 2011. The Visible Camera : Hand-Held Camera Movement and Cinematograpohic. Spectrums, Spring, pp.34–40.

Andersen, Thomas Hæstbæk, 2017. Metafunctions. Centre for Multimodal Communication. Available at: <https://multimodalkeyterms.wordpress.com/metafunctions/> [Accessed 21 May 2022].

Anneliene, A.A.L.F.M., Stupar-Rutenfrans, S., Bastiaens, O.S.P. and Van Gisbergen, M.M.S., 2015. Observe or participate: The effect of point-of-view on presence and enjoyment in 360 degree movies for head mounted displays. CEUR Workshop Proceedings, 1528.

Anon. 2018. Lexical Density. [online] Analyze My Writing. Available at: <https://www.analyzemywriting.com/lexical_density.html> [Accessed 5 February 2022].

Anon. 2022a. BBC News - YouTube. [online] Available at: <https://www.youtube.com/> [Accessed 20 February 2022].

Anon. 2022. Plan International UK - YouTube. [online] Available at: <https://www.youtube.com/> [Accessed 20 February 2022].

Anon. 2022b. Windows Tool for Speech Analysis. [online] UCL PSYCHOLOGY AND LANGUAGE SCIENCES. Available at: <https://www.phon.ucl.ac.uk/resource/sfs/wasp/> [Accessed 31 January 2022].

Arctic 360. 2016. Directed by B. Gustavsson. ARTE. Available at: <https://www.theguardian.com/environment/ng-interactive/2016/dec/22/arctic-360-take-a-tour-without-doing-damage>.

Arora, G. and Palitz, A., 2016. my mother's wing. [online] WITHIN. Available at: <http://with.in/watch/my-mothers-wing/>.

Arrivée d'un train (à la Ciotat). 1895. [35mm print] Directed by L. Lumiére. Available at: <https://www.moma.org/learn/moma_learning/louis-lumiere-arrivee-dun-train-a-la-ciotat-arrival-of-a-train-at-la-ciotat-1895/>.

Austin, J.L., 1975. How to Do Things with Words. Harvard University Press.

Azungah, T., 2018. Qualitative research: deductive and inductive approaches to data analysis. Qualitative Research Journal, 18(4), pp.383–400. https://doi.org/10.1108/QRJ-D-18-00035.

Baldry, A. and Thibault, P.J., 2006. Multimodal Transcription and Text Analysis: A Multimedia Toolkit and Coursebook. Equinox textbooks and surveys in Linguistics. [online] Equinox Publishing Limited. Available at: <https://books.google.ie/books?id=w9BpQgAACAAJ>.

Barnard, D., 2019. Degrees of Freedom (DoF): 3-DoF vs 6-DoF for VR Headset Selection. [online] Virtual Speech. Available at: <https://virtualspeech.com/blog/degrees-of-freedom-vr> [Accessed 23 April 2022].

Barthes, R., 1964. Rhetoric of the Image. na.

Bateman, J. and Schmidt, K.H., 2013. Multimodal Film Analysis: How Films Mean. Routledge Studies in Multimodality. [online] Taylor & Francis. Available at: <https://books.google.ie/books?id=_yLDrxnMmeYC>.

Bateman, J.A. and Wildfeuer, J., 2014. A multimodal discourse theory of visual narrative. Journal of Pragmatics, 74, pp.180–208. https://doi.org/10.1016/j.pragma.2014.10.001.

Becker, H., 1990. Qualitative inquiry in education: The continuing debate. In: Generalizing from case studies. Teachers College Press New York.pp.233–242.

Bertrand, P., Guegan, J., Robieux, L., McCall, C.A. and Zenasni, F., 2018. Learning Empathy Through Virtual Reality: Multiple Strategies for Training Empathy-Related Abilities Using Body Ownership Illusions in Embodied Virtual Reality. Frontiers in Robotics and AI, [online] 5. https://doi.org/10.3389/frobt.2018.00026.

Bevan, C. and Green, D., 2017. A Mediography Of Virtual Reality Non- Fiction : Insights And Future Directions. pp.161–166.

Bevan, C. and Green, D., 2018. VR Nonfiction: A Mediography. [online] Available at: <http://vrdocumentaryencounters.co.uk/vrmediography/vrmediography/listing/> [Accessed 11 October 2019].

Bezemer, J., 2014. How to transcribe multimodal interaction? A case study.

Bigelow, K., 2017. The Protectors. [online] Company 3 [us]. Available at: <http://www.imdb.com/title/tt6075918/?ref_=ttpl_pl_tt>.

Biocca, F., 2002. The evolution of interactive media: Toward. Being There" in Nonlinear.

Bolter, J. D. Grusin, R., 2000. Remediation: Understanding New Media. London; Cambridge, Mass.: MIT Press.

Bolter, J.D. and Grusin, R., 1999. Remediation: understanding new media. Cambridge, Mass: MIT Press.

Bouchard, S., St-Jacques, J., Robillard, G. and Renaud, P., 2008. Anxiety Increases the Feeling of Presence in Virtual Reality. Presence: Teleoperators and Virtual Environments, 17(4), pp.376–391. https://doi.org/10.1162/pres.17.4.376.

Bowcher, W.L. and Liang, J.Y. eds., 2016. Society in Language, Language in Society. [online] London: Palgrave Macmillan UK. https://doi.org/10.1057/9781137402868.

Branigan, E., 1984. Point of view in the cinema: a theory of narration and subjectivity in classical film. Approaches to semiotics. Berlin ; New York: Mouton.

Brillhart, J., 2018a. In the Blink of a Mind — Attention. [online] Medium. Available at: <https://medium.com/the-language-of-vr/in-the-blink-of-a-mind-attention-1fdff60fa045> [Accessed 7 July 2021].

Brillhart, J., 2018b. The Language Of VR. [online] Available at: <https://medium.com/the-language-of-vr> [Accessed 20 July 2018].

Brine, K.G., 2020. The Art of Cinematic Storytelling: A Visual Guide to Planning Shots, Cuts, and Transitions. Oxford University Press.

Broomfield, N., 1998. Kurt And Courtney. [online] Optimum. Available at: <https://www.imdb.com/title/tt0138563/?ref_=nv_sr_1>.

Brown, A., Sheikh, A., Evans, M. and Watson, Z., 2016a. Directing attention in 360-degree video. In: IBC 2016 Conference. [online] IBC 2016 Conference. Amsterdam, Netherlands: Institution of Engineering and Technology.p.29 (9 .)-29 (9 .). https://doi.org/10.1049/ibc.2016.0029.

Brown, A., Sheikh, A., Evans, M. and Watson, Z., 2016b. Directing attention in 360-degree video. p.29 (9 .)-29 (9 .). https://doi.org/10.1049/ibc.2016.0029.

Brown, B. and Safari, an O.M.C., 2016. Cinematography: Theory and Practice, 3rd Edition. [online]

Available at: <https://www.safaribooksonline.com/complete/auth0oauth2/&state=/library/view//9781 317359265/?ar> [Accessed 22 June 2021].

Brown, T., 2013. *Breaking the Fourth Wall*. Edinburgh University Press.

Bruner, J., 2009. *Actual Minds, Possible Worlds*. revised ed. The Jerusalem-Harvard Lectures. [online] Harvard University Press. Available at: <https://books.google.ie/books?id=YNuBf6W2rt0C>.

Bruzzi, S., 2000. *New documentary: a critical introduction*.

Bucher, J., 2017. *Storytelling for Virtual Reality: Methods and Principles for Crafting Immersive Narratives*. [online] Taylor & Francis. Available at: <https://books.google.ie/books?id=hkgrDwAAQBAJ>.

Caldwell, K., 2015. *Nick Broomfield reveals working with Courtney Love was a 'living hell' | Daily Mail Online*. [online] The Daily Mail Online. Available at: <https://www.dailymail.co.uk/tvshowbiz/article-3054691/Kurt-Courtney-director-Nick-Broomfield-reveals-working-Love-1998-documentary-living-hell.html> [Accessed 18 May 2019].

Cartier-Bresson, H., 1953. *Pablo Picasso*. [Gelatin silver Photographic print].

Castaing-Taylor, L. and Paravel, V., 2012. *Leviathan*. [online] Dogwoof. Available at: <http://www.imdb.com/title/tt2332522/?ref_=nv_sr_5>.

Castello, E., 2008. *Text Complexity and Reading Comprehension Tests*. Peter Lang.

Castells, A.G., 2011. *The Interactive Documentary. Definition Proposal and Characterization of the New Emerging Genre*. Hipertext.net.

Cheong, Y.Y., 2004. The construal of Ideational meaning in print advertisements. In: K.L. O'Halloran, ed. *Multimodal discourse analysis: systemic-functional perspectives, Open linguistics series*. London ; New York: Continuum.

Cohen, J., 2001. Defining Identification: A Theoretical Look at the Identification of Audiences With Media Characters. *Mass Communication & Society*, 4(3), pp.245–264. https://doi.org/10.1207/S15327825MCS0403.

Cone, J., 2015. *Uncovering the Grammar of VR with Saschka Unseld*. [online] Motionographer. Available at: <https://motionographer.com/2015/10/01/uncovering-the-grammar-of-vr-with-saschka-unseld/> [Accessed 21 March 2022].

Congo VR: A Troubled Past. 2018. Directed by P. Harper. BBC. Available at: <http://vrdocumentaryencounters.co.uk/vrmediography/detail/476/> [Accessed 17 January 2020].

Congo VR: A Troubled Past - BBC News. 2019. Directed by YouTube. Available at: <https://www.youtube.com/watch?v=Mwns3S-kSSA> [Accessed 17 May 2022].

Corner, J., 2000. What can we say about 'documentary'? *Media Culture and Society*, [online] Vol 22(Issue 5). Available at: <http://journals.sagepub.com/doi/pdf/10.1177/016344300022005009>.

Cowie, E., 2011. *Recording reality, desiring the real*. [online] Minneapolis: University of Minnesota Press. Available at: <https://books.google.ie/books?id=YBWtb8zbNBAC&printsec=frontcover&dq=elizabeth+co wie&hl=en&sa=X&ved=2ahUKEwi-j_Oclun2AhWCi1wKHXPnAhkQuwV6BAgFEAc#v=snippet&q=a%20look&f=false>.

Coxon, M., Kelly, N. and Page, S., 2016. *Individual differences in virtual reality: Are spatial presence*

and spatial ability linked? *Virtual Reality*, 20(4), pp.203–212. https://doi.org/10.1007/s10055-016-0292-x.

Creswell, J.W., 2014. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches.* [online] SAGE Publications. Available at: <https://books.google.ie/books?id=4uB76IC_pOQC>.

Davidson, C., 2009. Transcription: Imperatives for Qualitative Research. *International Journal of Qualitative Methods*, 8(2), pp.35–52. https://doi.org/10.1177/160940690900800206.

Dinesh, N., 2016. *Memos from a Theatre Lab: Exploring what immersive theatre 'does'.* [online] Taylor & Francis. Available at: <https://books.google.ie/books?id=0VyuDQAAQBAJ>.

Dobson, S., 2009. Remediation. Understanding New Media - Revisiting a Classic. *Seminar.net - International Journal of Media, Technology & Lifelong Learning*, 5(2), pp.1–9.

Dockers. 1999. Directed by B. Anderson. Available at: <https://www.imdb.com/title/tt0206691/> [Accessed 5 May 2019].

Domingo, M., 2011. Analyzing layering in textual design: A multimodal approach for examining cultural, linguistic, and social migrations in digital video. *International Journal of Social Research Methodology*, 14(3), pp.219–230. https://doi.org/10.1080/13645579.2011.563619.

Dooley, K., 2017a. Storytelling with virtual reality in 360-degrees: a new screen grammar. *Studies in Australasian Cinema*, 11(3), pp.161–171. https://doi.org/10.1080/17503175.2017.1387357.

Dooley, K., 2017b. Storytelling with virtual reality in 360-degrees: a new screen grammar. *Studies in Australasian Cinema*, 11(3), pp.161–171. https://doi.org/10.1080/17503175.2017.1387357.

Dooley, K., 2020. A question of proximity: exploring a new screen grammar for 360-degree cinematic virtual reality. *Media Practice and Education*, 21(2), pp.81–96. https://doi.org/10.1080/25741136.2019.1641005.

Dredge, S., 2015. Virtual reality documentaries 'take the middle man out of journalism'. *The Guardian*. [online] 29 Jan. Available at: <https://www.theguardian.com/technology/2015/jan/29/virtual-reality-documentary-middle-man-journalism-chris-milk-film> [Accessed 18 May 2019].

Eggins, S., 2004. *Introduction to Systemic Functional Linguistics: 2nd Edition. An Introduction to Systemic Functional Linguistics.* [online] Bloomsbury Academic. Available at: <https://books.google.ie/books?id=sS7UXugIIg8C>.

Eisenhardt, K.M., 1989. Building Theories from Case Study Research. *Academy of Management Review*.

Elmezeny, A., Edenhofer, N. and Wimmer, J., 2018. Immersive Storytelling in 360-Degree Videos: An Analysis of Interplay Between Narrative and Technical Immersion. 11(1), p.15.

EPSRC, 2019. *Virtual Realities - Immersive Documentary Encounters.* [online] Available at: <https://gow.epsrc.ukri.org/NGBOViewGrant.aspx?GrantRef=EP/P025595/1> [Accessed 15 November 2019].

Extreme Love - Dementia. 2012. Directed by L. Theroux. Available at: <https://www.imdb.com/title/tt5111874/?ref_=nm_knf_t2> [Accessed 5 July 2019].

Fahrenheit 9/11. 2004. Directed by M. Moore. Fellowship Adventure Group, Dog Eat Dog Films, Miramax.

Feng, D. and O'Halloran, K.L., 2013. The visual representation of metaphor: A social semiotic approach. *Review of Cognitive Linguistics. Published under the auspices of the Spanish*

Cognitive Linguistics Association, 11(2), pp.320–335. https://doi.org/10.1075/rcl.11.2.07fen.

Fortner, R.S. and Fackler, P.M., 2014. The Handbook of Media and Mass Communication Theory. Handbooks in Communication and Media. [online] Wiley. Available at: <https://books.google.ie/books?id=ThkRAwAAQBAJ>.

Frischen, A., 2007. Gaze Cueing of Attention: p.57.

Gaudenzi, S., 2013. The Living Documentary: from representing reality to co-creating reality in digital interactive documentary. (January).

Gaudreault, A. and Marion, P., 2002. Genealogy Model for the of Media. Convergence.

Geiger, J., 2008. The camera and man The Camera and Man Colonialism , Masculinity and Documentary Fiction. 8822(1998), pp.3–17. https://doi.org/10.1080/09528829808576714.

George, G.H.S.P. of I.R.A.L., George, A.L., Bennett, A., Bennett, P. of E.A., Lynn-Jones, S.M. and Miller, S.E., 2005. Case Studies and Theory Development in the Social Sciences. MIT Press.

Gibson, J.J., 1979. The Ecological Approach To Visual Perception. [online] Taylor & Francis. Available at: <https://books.google.ie/books?id=yv_9hU_26KEC>.

Goffman, E., 1959. The Presentation of Self in Everyday Life. Anchor books. [online] Anchor Books/Doubleday. Available at: <https://books.google.ie/books?id=1G89PgAACAAJ>.

Goffman, E., 1979. Footing. Semiotica, 25(1–2), pp.1–30. https://doi.org/10.1515/semi.1979.25.1-2.1.

Gomm, R., Hammersley, M. and Foster, P., 2000. Case Study Method: Key Issues, Key Texts. [online] SAGE Publications. Available at: <https://books.google.ie/books?id=c1VOCAGy86EC>.

Gregory, S., 2016. Immersive Witnessing: From Empathy and Outrage to Action. [online] WITNESS Blog. Available at: <https://blog.witness.org/2016/08/immersive-witnessing-from-empathy-and-outrage-to-action/> [Accessed 31 December 2021].

Grey Gardens. 1975. Directed by E. Hovde. and Directed by A. Maylses. Available at: <http://www.imdb.com/title/tt0073076/> [Accessed 18 May 2019].

Gunning, T., 1986. The Cinema of Attraction. Early Film, Its Spectator and the Avant-Garde. [online] Available at: <http://filmlexikon.uni-kiel.de/index.php?action=lexikon&tag=det&id=1364>.

Haedicke, S.C., 2002. The Politics of Participation: Un Voyage Pas Comme Les Autres Sur Les Chemins De L'Exil. Theatre topics, 12(2), pp.99–118.

Hall, E.T., 1966. The Hidden Dimension. [online] Anchor Books. Available at: <https://books.google.ie/books?id=HlmqAAAACAAJ>.

Hall, J., 1991. Society for Cinema & Media Studies Realism as a Style in Cinema Verite : A Critical Analysis of " Primary " Author (s): Jeanne Hall Published by : University of Texas Press on behalf of the Society for Cinema & Media Studies Stable URL : http://www.jsto. 30(4), pp.24–50.

Halliday, M.A.K., 1978. Language as Social Semiotic: The Social Interpretation of Language and Meaning. Open University set book. [online] Edward Arnold. Available at: <https://books.google.ie/books?id=SjVxAAAAIAAJ>.

Halliday, M.A.K. and Hasan, R., 1976. Cohesion in English. Routledge.

Halliday, M.A.K. and Hasan, R., 1985. Language, Context, and Text: Aspects of Language in a Social-semiotic.

Halliday, M.A.K. and Matthiessen, C.M.I.M., 1999. Construing experience through meaning: a

language-based approach to cognition. Open linguistics series. London; New York: Continuum.

Halliday, M.A.K. and Matthiessen, C.M.I.M., 2013. Halliday's Introduction to Functional Grammar. Routledge.

Henley, P., 2010. The Adventure of the Real: Jean Rouch and the Craft of Ethnographic Cinema. University of Chicago Press.

Hjelmslev, L., 1961. Prolegomena to a Theory of Language. University of Wisconsin Press.

Hood, S. and Hao, J., 2021. Grounded learning: Telling and showing in the language and paralanguage of a science lecture. In: Teaching Science. Routledge.

Hospital. 1970. Directed by F. Wiseman. Available at: <http://www.imdb.com/title/tt0065853/> [Accessed 18 May 2019].

Hull, G.A. and Nelson, M.E., 2005. Locating the semiotic power of multimodality. https://doi.org/10.1177/0741088304274170.

Industrial Britain. 1931. Directed by R.J. Flaherty. Empire Marketing Board Film Unit.

Irwin, W., 2002. The Matrix and Philosophy: Welcome to the Desert of the Real. Popular culture and philosophy. [online] Open Court. Available at: <https://books.google.ie/books?id=GSE5qlwGzCAC>.

Itti, L., 2005. Quantifying the contribution of low-level saliency to human eye movements in dynamic scenes. Visual Cognition, 12(6), pp.1093–1123. https://doi.org/10.1080/13506280444000661.

Jaffe, A., 2000. Introduction: Non-standard orthography and non-standard speech. Journal of Sociolinguistics, 4(4), pp.497–513. https://doi.org/10.1111/1467-9481.00127.

Jaremko-Greenwold, A., 2013. Véréna Paravel and Lucien Castaing-Taylor by Anya Jaremko-Greenwold. [online] Available at: <https://bombmagazine.org/articles/véréna-paravel-and-lucien-castaing-taylor/>.

Jerald, J., 2015. The VR Book: Human-Centered Design for Virtual Reality. ACM Books. [online] Association for Computing Machinery and Morgan & Claypool Publishers. Available at: <https://books.google.ie/books?id=ZEBiDwAAQBAJ>.

Jewitt, C., 2016. The Routledge Handbook of Multimodal Analysis. Routledge.

Kahana, J. and Musser, C., 2016. The Documentary Film Reader: History, Theory, Criticism. [online] Oxford University Press. Available at: <https://books.google.ie/books?id=xVU0CwAAQBAJ>.

Keating, P., 2014. Cinematography. Rutgers University Press.

Kendon, A., 1967. Some functions of gaze-direction in social interaction. Acta Psychologica, 26, pp.22–63. https://doi.org/10.1016/0001-6918(67)90005-4.

Kendon, A., 1990. Conducting Interaction: Patterns of Behavior in Focused Encounters. CUP Archive.

Kendon, A., 2010. Spacing and Orientation in Co-present Interaction. In: A. Esposito, N. Campbell, C. Vogel, A. Hussain and A. Nijholt, eds. Development of Multimodal Interfaces: Active Listening and Synchrony: Second COST 2102 International Training School, Dublin, Ireland, March 23-27, 2009, Revised Selected Papers, Lecture Notes in Computer Science. [online] Berlin, Heidelberg: Springer.pp.1–15. https://doi.org/10.1007/978-3-642-12397-9_1.

Keskinen, T., Makela, V., Kallioniemi, P., Hakulinen, J., Karhu, J., Ronkainen, K., Makela, J. and Turunen, M., 2019. The Effect of Camera Height, Actor Behavior, and Viewer Position on the User Experience of 360° Videos. In: 2019 IEEE Conference on Virtual Reality and 3D User

Interfaces (VR). [online] 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR). Osaka, Japan: IEEE.pp.423–430. https://doi.org/10.1109/VR.2019.8797843.

Kjær, T., Lillelund, C.B., Moth-Poulsen, M., Nilsson, N.C., Nordahl, R. and Serafin, S., 2017. Can you cut it?: an exploration of the effects of editing in cinematic virtual reality. In: Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology. [online] VRST '17: 23rd ACM Symposium on Virtual Reality Software and Technology. Gothenburg Sweden: ACM.pp.1–4. https://doi.org/10.1145/3139131.3139166.

Kool, H., 2016a. The Ethics of Immersive Journalism : A rhetorical analysis of news storytelling with virtual reality technology. Intersect, 9(3), pp.1–11.

Kool, H., 2016b. The Ethics of Immersive Journalism : A rhetorical analysis of news storytelling with virtual reality technology. Intersect, 9(3), pp.1–11.

Kress, G., 2001. Multimodal Teaching and Learning: The Rhetorics of the Science Classroom. Advances in Applied Linguistics. [online] Bloomsbury Academic. Available at: <https://books.google.ie/books?id=YnfZm5z37WMC>.

Kress, G., 2009. Multimodality: A social semiotic approach to contemporary communication. Routledge.

Kress, G., Jewitt, C., Ogborn, J. and Charalampos, T., 2001. Multimodal Teaching and Learning: The Rhetorics of the Science Classroom. Bloomsbury Classics in Linguistics. [online] Bloomsbury Publishing. Available at: <https://books.google.ie/books?id=_gxCBAAAQBAJ>.

Kress, G.R. and van Leeuwen, T., 1996. Reading Images: The Grammar of Visual Design. 2nd ed.

Kress, G.R. and van Leeuwen, T., 2006. Reading Images: The Grammar of Visual Design. 2nd ed.

Kress, G.R. and Van Leeuwen, T., 2001. Multimodal Discourse: The Modes and Media of Contemporary Communication.

La Seine a rencontré Paris. 1957. Directed by J. Ivens. Garance.

Labov, W., 2013. The Language of Life and Death: The Transformation of Experience in Oral Narrative. Cambridge University Press.

Lajeunesse, F., 2017. Cirque du Soleil: Dreams of 'O' VR.

Lam, M., 2016. Interfacing Field with Tenor: Hasan's Notion of Personal Distance. In: W.L. Bowcher and J.Y. Liang, eds. Society in Language, Language in Society. [online] London: Palgrave Macmillan UK.pp.206–226. https://doi.org/10.1057/9781137402868.

Lancaster, K., 2019. Basic Cinematography: A Creative Guide to Visual Storytelling. Routledge.

Larsson, P., Västfjäll, D. and Kleiner, M., 2001. The Actor-Observer Effect in Virtual Reality Presentations. CyberPsychology & Behavior, 4(2), pp.239–246. https://doi.org/10.1089/109493101300117929.

Laurie M. Wilcox, Robert S. Allison, Samuel Elfassy, C.G., 2006. Personal Space in Virtual Reality. ACM Transactions on Applied Perception (TAP), 3(4), pp.412–428.

van Leeuwen, T., 1991. Conjunctive structure in documentary film and television. Continuum, 5(1), pp.76–114. https://doi.org/10.1080/10304319109388216.

van Leeuwen, T., 2005. Introducing Social Semiotics. Psychology Press.

Lemke, J.L., 2004. METAMEDIA LITERACY: TRANSFORMING MEANINGS AND MEDIA. p.21.

Lemke, J.L., 2005. Multiplying meaning Visual and verbal semiotics in scientific text. In: J.R. Martin, ed. Reading Science: Critical and Functional Perspectives on Discourses of Science, 1st ed. [online] Routledge. https://doi.org/10.4324/9780203982327.

Lemke, J.L., 2006. Opening Up Closure: Semiotics Across Scales. Annals of the New York Academy of Sciences, 901(1), pp.100–111. https://doi.org/10.1111/j.1749-6632.2000.tb06269.x.

Lim Fei, V., 2013. The Visual Semantics Stratum: Making Meaning in Sequential Images. In: T.D. Royce and W. Bowcher, eds. New Directions in the Analysis of Multimodal Discourse, 0 ed. [online] Routledge.pp.200–218. https://doi.org/10.4324/9780203357774-11.

Lim, F.V., 2004. Developing an integrative multi-semiotic model. In: K.L.. O'Halloran, ed. Multimodal Discourse Analysis: Systemic Functional Perspectives. A&C Black.pp.220–246.

Lister, M., 2009. New media : a critical introduction. 2nd ed. Milton Park, Abingdon, Oxon ; New York, N.Y.: Routledge.

Liu, Y. and O'Halloran, K.L., 2009. Intersemiotic Texture: analyzing cohesive devices between language and images. Social Semiotics, 19(4), pp.367–388. https://doi.org/10.1080/10350330903361059.

Love, N., 2012. The linguistic thought of J.R. Firth. Stellenbosch Papers in Linguistics, [online] 15(0). https://doi.org/10.5774/15-0-96.

M Hammersley and R Gomm, 2000. Case Study and Generalisation. In: P. Foster, M. Hammersley and R. Gomm, eds. Case study method. pp.98–116.

Machon, J., 2009. (Syn)aesthetics: Redefining Visceral Performance. Springer.

Mackenzie, J., 2016. All You Need to Know about 360 VR Photos. Vantage. Available at: <https://medium.com/vantage/all-you-need-to-know-about-360-vr-photos-d180aa25d8de> [Accessed 3 June 2019].

MacMillan, K., 2005. More Than Just Coding? Evaluating CAQDAS in a Discourse Analysis of News Texts. Forum Qualitative Sozialforschung / Forum: Qualitative Social Research, [online] 6(3). https://doi.org/10.17169/fqs-6.3.28.

Malick, T., 2018. Together. [online] Available at: <https://www.youtube.com/watch?v=pEzGqeNuHrg>.

Mamie's Dream. 2016. Directed by M. Matheson. Plan International UK. Available at: <http://vrdocumentaryencounters.co.uk/vrmediography/detail/81/> [Accessed 17 January 2020].

Mamie's Dream, her full story: in virtual reality. 2016. Plan International UK Available at: <https://www.youtube.com/watch?v=EnDyI4kN2zU> [Accessed 20 February 2022].

Man of Aran. 1934. Directed by R.J. Flaherty. Gainsborough Pictures.

Marquis, E., 2014. Conceptualizing Documentary Performance. Studies in Documentary Film, 7(1), pp.45–60. https://doi.org/10.1386/sdf.7.1.45.

Marshall, P., Rogers, Y. and Pantidi, N., 2011. Using F-formations to analyse spatial patterns of interaction in physical environments. In: Proceedings of the ACM 2011 conference on Computer supported cooperative work - CSCW '11. [online] the ACM 2011 conference. Hangzhou, China: ACM Press.p.445. https://doi.org/10.1145/1958824.1958893.

Martin, J.R., 1984. Language, Register and Genre. In: F. Christie, R.M. Bunbury and S. Dawkins, eds. Language Studies: Children Writing. Reader, ECT418 language studies. [online] Deakin University. Available at: <https://books.google.ie/books?id=pMBHKgAACAAJ>.

Martin, J.R., 1992. English text: system and structure. Philadelphia: Benjamins.

Martin, J.R., 2002. 3. Meaning Beyond the Clause: Sfl Perspectives. Annual Review of Applied Linguistics, 22, pp.52–74. http://dx.doi.org/10.1017/S026719050200003X.

Martin, J.R., 2005. *Reading Science: Critical and Functional Perspectives on Discourses of Science*. 1st ed. [online] Routledge. https://doi.org/10.4324/9780203982327.

Martin, J.R. and Rose, D., 2007. *Working with Discourse Second Edition: Meaning Beyond the Clause*. Bloomsbury Academic.

Martin, J.R. and Rose, D., 2008. *Genre Relations: Mapping Culture. Equinox textbooks and surveys in linguistics*. [online] Equinox Pub. Available at: <https://books.google.ie/books?id=NbIgPwAACAAJ>.

Martin, J.R. and White, P.R.R., 2005. *The language of evaluation: appraisal in English*. New York: Palgrave Macmillan.

Martinec, R., 1998. *Cohesion in action. Semiotica*, [online] 120(1–2). https://doi.org/10.1515/semi.1998.120.1-2.161.

Martinec, R., 2000. *Rhythm in Multimodal Texts. Leonardo*, 33(4), pp.289–297. https://doi.org/10.1162/002409400552676.

Martinec, R., 2001. *Interpersonal resources in action. Semiotica*, 2001(135), pp.117–145. https://doi.org/10.1515/semi.2001.056.

Martinec, R. and Salway, A., 2005. *A system for image–text relations in new (and old) media. Visual communication*, 4(3), pp.337–371.

Matthiessen, C.M., 2015. *Register in the round: registerial cartography. Functional Linguistics*, 2(1), p.9. https://doi.org/10.1186/s40554-015-0015-8.

Matthiessen, C.M.I.M., 2009. *Multisemiosis and Context- Based Register Typology: Registerial Variation in the Complementarity of Semiotic Systems. In: E. Ventola and A.J.M. Guijarro, eds. The World Told and the World Shown: Multisemiotic Issues*. Springer.pp.11–38.

Matthiessen, C.M.I.M., 2019. *Register in Systemic Functional Linguistics*. Register Studies*, 1(1), pp.10–41. https://doi.org/10.1075/rs.18010.mat.

Matwick, K. and Matwick, K., 2014. *Storytelling and synthetic personalization in television cooking shows. Journal of Pragmatics*, 71, pp.151–159. https://doi.org/10.1016/j.pragma.2014.08.005.

McRoberts, J., 2017. *Are we there yet? Media content and sense of presence in non-fiction virtual reality. Studies in Documentary Film*, pp.1–18. https://doi.org/10.1080/17503280.2017.1344924.

Mechura, M., 2005. *A practical guide for functional text analysis: Analyzing English texts for field, mode, tenor and communicative effectiveness*. 1, p.2007.

Měchura, M.B., 2005. *A PRACTICAL GUIDE FOR FUNCTIONAL TEXT ANALYSIS*.

Menand, L., 2009. *Some Frames for Goffman. Social Psychology Quarterly*, 72(4), pp.296–299. https://doi.org/10.1177/019027250907200403.

Metz, C., 1974. *On the Impression of Reality in the Cinema. Film Language : A Semiotics of the Cinema*, p.268.

Miles, M.B. and Huberman, A.M., 1994. *Qualitative Data Analysis: An Expanded Sourcebook*. SAGE.

Milk, C., 2016. *The birth of virtual reality as an art form*. [online] Ted.com. Available at: <https://www.ted.com/talks/chris_milk_the_birth_of_virtual_reality_as_an_art_form>.

Milk, C. and Arora, G., 2015a. *clouds over sidra*.

Milk, C. and Arora, G., 2015b. *Waves Of Grace*. [online] Available at: <http://with.in/watch/waves-of-grace/>.

MIT, 2019. MIT – Docubase. [Academic Educational] MIT - Docubase. Available at: <https://docubase.mit.edu> [Accessed 22 November 2019].

Moana. 1926. Directed by F.H. Flaherty. and Directed by R.J. Flaherty. Robert Flaherty Productions Inc.

Nanook of the North. 1922. Directed by R.J. Flaherty. Les Frères Revillon.

Nash, K., 2017. Virtual reality witness: exploring the ethics of mediated presence. Studies in Documentary Film, 0(0), pp.1–13. https://doi.org/10.1080/17503280.2017.1340796.

New York Times, 2016. Immersive (AR/VR). [online] The New York Times. Available at: <https://www.nytimes.com/spotlight/augmented-reality> [Accessed 24 January 2020].

Nichols, B., 1983. The Voice of Documentary. Film Quarterly-University of California Press, 36(3), pp.17–30.

Nichols, B., 2001. Introduction to documentary. [online] Available at: <http://books.google.de/books?id=tFLMar9GsjQC>.

Nichols, B., 2010. Introduction to documentary. 2nd ed. Palo Alto, Calif.: Indiana University Press.

Night Fall. 2016. Directed by J. Samhoud. and Directed by M. Korver. Available at: <vrdocumentaryencounters.co.uk/vrmediography/detail/522/> [Accessed 17 January 2020].

Night Mail. 1936. Directed by H. Watt. and Directed by B. Wright. Available at: <https://www.imdb.com/title/tt0250616/?ref_=nv_sr_1?ref_=nv_sr_1> [Accessed 5 May 2019].

O'Connell, D.C. and Kowal, S., 1994. Some current transcription systems for spoken discourse: A critical analysis. Pragmatics, 4(1), pp.81–107. https://doi.org/10.1075/prag.4.1.04con.

O'Donnell, M., 2020. Dynamic modelling of context: Field, Tenor and Mode revisited. Lingua, p.102952. https://doi.org/10.1016/j.lingua.2020.102952.

O'Grady, G., 2017. Intonation and systemic functional linguistics The way forward. In: T. Bartlett, ed. The Routledge Handbook of Systemic Functional Linguistics. Taylor & Francis.pp.146–162.

O'Halloran, K., 2004. Multimodal Discourse Analysis: Systemic Functional Perspectives. A&C Black.

O'Halloran, K.L.., 2008a. Systemic functional-multimodal discourse analysis (SF-MDA): constructing ideational meaning using language and visual imagery. Visual Communication, 7(4), pp.443–475. https://doi.org/10.1177/1470357208096210.

O'Halloran, K.L.., 2008b. Systemic functional-multimodal discourse analysis (SF-MDA): constructing ideational meaning using language and visual imagery. Visual Communication, 7(4), pp.443–475. https://doi.org/10.1177/1470357208096210.

O'Halloran, K.L., Tan, S. and Wignell, P., 2019. SFL and Multimodal Discourse Analysis. In: G. Thompson, W.L. Bowcher and L. Fontaine, eds. The Cambridge Handbook of Systemic Functional Linguistics, 1st ed. [online] Cambridge University Press.pp.433–461. https://doi.org/10.1017/9781316337936.019.

O'Neill, K., 2015. The UN's First Virtual Reality Documentary Puts You Inside a Syrian Refugee Camp. Vice. Available at: <https://www.vice.com/en/article/jpv5ng/virtual-reality-doc-puts-you-inside-a-syrian-refugee-camp> [Accessed 1 May 2022].

O'Toole, M., 1994. The Language of Displayed Art. [online] Fairleigh Dickinson University Press. Available at: <https://books.google.ie/books?id=U67PhHo4gfwC>.

Passmore, P.J., Glancy, M., Philpot, A. and Fields, B., 2017. 360 CINEMATIC LITERACY: A CASE STUDY. p.9.

Paulus, T.M. and Lester, J.N., 2016. ATLAS.ti for conversation and discourse analysis studies. *International Journal of Social Research Methodology*, 19(4), pp.405–428. https://doi.org/10.1080/13645579.2015.1021949.

Pavsek, C., 2015. Leviathan and the experience of sensory ethnography. *Visual Anthropology Review*, 31(1), pp.4–11. https://doi.org/10.1111/var.12056.

Plan International, 2022. About - Child Rights - Equality for Girls. [online] *Plan International Ireland*. Available at: <https://www.plan.ie/about-plan/what-we-do/> [Accessed 20 February 2022].

Planet Earth II: Deserts. 2016. Available at: <http://vrdocumentaryencounters.co.uk/vrmediography/detail/492/> [Accessed 22 January 2020].

Plantinga, C.R., 1997. *Rhetoric and Representation in Nonfiction Film.* Cambridge, UK: Cambridge University Press.

Poerio, G.L., Blakey, E., Hostler, T.J. and Veltri, T., 2018. More than a feeling: Autonomous sensory meridian response (ASMR) is characterized by reliable changes in affect and physiology. *PLOS ONE*, 13(6), p.e0196645. https://doi.org/10.1371/journal.pone.0196645.

Power and the Land. 1940. Directed by J. Ivens. Rural Electrification Administration, U.S. Department of Agriculture, United States Film Service.

Poynton, C., 1985. Language and Gender: Making the Difference. [online] Available at: <https://books.google.ie/books?id=og4cAQAAIAAJ&q=cate+poynton+language+and+gender&dq=cate+poynton+language+and+gender&hl=en&sa=X&redir_esc=y> [Accessed 28 November 2021].

Poynton, C.M., 1991. *Address and the semiotics of social relations.*

Primary. 1960. Directed by R. Drew. Available at: <http://www.imdb.com/title/tt0054205/> [Accessed 18 May 2019].

Renov, M., 1993. *Theorizing documentary.* New York: Routledge.

Rick, B., 2016. Film Scoring 101 - Atonality and Clusters - YouTube. [online] *YouTube.com*. Available at: <https://www.youtube.com/watch?v=qmCeA96lJ4s&t=196s> [Accessed 30 May 2019].

Riva, G., Wiederhold, B.K. and Mantovani, F., 2019. Neuroscience of Virtual Reality: From Virtual Exposure to Embodied Medicine. *Cyberpsychology, Behavior, and Social Networking*, 22(1), pp.82–96. https://doi.org/10.1089/cyber.2017.29099.gri.

Rogers, A.B., 2015. Studies in Documentary Film The crisis of performance and performance of crisis : the powers of the false in Grey Gardens (1976). 3280(September), pp.114–126. https://doi.org/10.1080/17503280.2015.1031565.

Rose, M., 2016. The empathy machine": hype, hope and immersion in non-fiction VR. [online] Available at: <https://idocs2016.dcrc.org.uk/abstracts/the-empathy-machine-hype-hope-and-immersion-in-non-fiction-vr/> [Accessed 19 June 2018].

Rothe, S. and Hussmann, H., 2019. Spaceline: A Concept for Interaction in Cinematic Virtual Reality. *Lecture Notes in Computer Science.* [online] Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-33894-7.

Rothe, S., Kegeles, B., Allary, M. and Hußmann, H., 2018. The impact of camera height in cinematic virtual reality. In: *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology.* [online] VRST '18: 24th ACM Symposium on Virtual Reality Software and Technology. Tokyo Japan: ACM.pp.1–2. https://doi.org/10.1145/3281505.3283383.

Rothman, W., 2004. *The 'I' of the Camera: Essays in Film Criticism, History, and Aesthetics.* Cambridge

Studies in Film. [online] Cambridge University Press. Available at: <https://books.google.ie/books?id=R92ziB67AMIC>.

Rouch, J., 1958. Moi Un Noir. [online] Available at: <https://www.imdb.com/title/tt0051942/>.

Rouse, R., 2016. Media of Attraction: A Media Archeology Approach to Panoramas, Kinematography, Mixed Reality and Beyond. Interactive Storytelling, [online] (9th International Conference on Interactive Digital Storytelling, ICIDS 2016, Los Angeles, CA, USA, November 15–18, 2016, Proceedings). Available at: <http://www.springer.com/us/book/9783319482781>.

Royce, T., 1998. Synergy on the page: Exploring intersemiotic complementarity in page-based multimodal text. JASFL Occasional papers, 1(1), pp.25–49.

Royce, T.D., 2007. Intersemiotic Complementarity: A Framework for Multimodal Discourse Analysis. In: New Directions in the Analysis of Multimodal Discourse. Routledge.

Royce, T.D., 2015. Intersemiotic Complementarity in Legal Cartoons: An Ideational Multimodal Analysis. International Journal for the Semiotics of Law - Revue internationale de Sémiotique juridique, 28(4), pp.719–744. https://doi.org/10.1007/s11196-015-9421-1.

Rubo, M. and Gamer, M., 2021. Stronger reactivity to social gaze in virtual reality compared to a classical laboratory environment. British Journal of Psychology, p.14.

Saldana, J., 2012. The Coding Manual for Qualitative Researchers. [online] SAGE Publications. Available at: <https://books.google.ie/books?id=V3tTG4jvgFkC>.

Salesman. 1969. Directed by A. Maylses. and Directed by D. Maysles. Available at: <http://www.imdb.com/title/tt0064921/> [Accessed 18 May 2019].

Samsung, 2015. Gear VR: How Samsung makes Virtual Reality a Reality. [online] Available at: <https://news.samsung.com/global/gear-vr-how-samsung-makes-virtual-reality-a-reality> [Accessed 30 March 2022].

Searching for Sugar Man. 2012. Directed by M. Bendjelloul. Red Box Films, Passion Pictures, Canfield Pictures.

Seawright, J. and Gerring, J., 2008. Case selection techniques in case study research: A menu of qualitative and quantitative options. Political research quarterly, 61(2), pp.294–308.

Sex: My British Job. 2013a. Directed by N. Broomfield. Lafayette Films. Available at: <http://www.imdb.com/title/tt3196598/?ref_=nv_sr_5>.

Sex: My British Job. 2013b. Nick Broomfield Available at: <https://www.youtube.com/watch?v=_wB8h-v6RLU> [Accessed 24 May 2022].

Shaw, S., 2008. Film Consciousness: From Phenomenology to Deleuze. Critical Explorations in Science Fiction and Fantasy Series. [online] McFarland & Company. Available at: <https://books.google.ie/books?id=dhIzBwAAQBAJ>.

Shore, S., 2014. Theme in Translation: A Systemic Functional Linguistic Perspective. International Journal of Comparative Literature and Translation Studies, [online] 2(4). https://doi.org/10.7575/aiac.ijclts.v.2n.4p.54.

Simons, H., 2009. Case Study Research in Practice. [online] SAGE Publications. Available at: <https://books.google.ie/books?id=7WHMNYEe434C>.

Slater, M., 2003. A Note on Presence Terminology. Emotion, 3, pp.1–5.

Slater, M., 2009. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. Philosophical Transactions Of The Royal Society, [online] 364(1535). Available at: <https://royalsocietypublishing.org/doi/10.1098/rstb.2009.0138> [Accessed 23 April

2022].

Smith, E., 2019. THE MAKING OF POWER FOR THE PARKINSONS. [History] powerforparkinsons.com. Available at: <http://www.powerforparkinsons.com/biptown.html> [Accessed 5 May 2019].

Sobchack, V., 1990. The active eye: A phenomenology of cinematic vision. Quarterly Review of Film and Video, 12(3), pp.21–36. https://doi.org/10.1080/10509209009361350.

Sobchack, V., 2004. Carnal Thoughts: Embodiment and Moving Image Culture. [online] University of California Press. Available at: <https://books.google.ie/books?id=IcnvecT955QC>.

Starman, A.B., 2013. The case study as a type of qualitative research. p.17.

Stenglin, M.K., 2009. Space odyssey: towards a social semiotic model of three-dimensional space. Visual Communication, 8(1), pp.35–64. https://doi.org/10.1177/1470357208099147.

Super Size Me. 2004. Directed by M. Spurlock., Directed by M. Spurlock., Directed by D. Isaacs. and Directed by C. Walker. The Con, Kathbur Pictures, Studio On Hudson.

Sutherland, E.A., 2012a. Staged Empathy: Empathy and Visual Perception in Virtual Reality Systems. p.107.

Sutherland, E.A., 2012b. Staged Empathy: Empathy and Visual Perception in Virtual Reality Systems Empathy and Visual Perception in Virtual Reality Systems.

Tan, S., O'Halloran, K.L. and Wignell, P., 2016. Multimodal research: Addressing the complexity of multimodal environments and the challenges for CALL. ReCALL, 28(3), pp.253–273. https://doi.org/10.1017/S0958344016000124.

Tench, P., 1996. The Intonation Systems of English. Bloomsbury Publishing.

Tench, P., 2020. 11. Intonation in semantic system networks. In: Approaches to Systemic Functional Grammar. p.23.

The Birds. 1963. Directed by A. Hitchcock., Directed by R. Taylor., Directed by T. Hedren. and Directed by J. Tandy. Alfred J. Hitchcock Productions.

The March of Time, 1935. RKO.

Thomas, D.R., 2006. A General Inductive Approach for Analyzing Qualitative Evaluation Data. American Journal of Evaluation, 27(2), pp.237–246. https://doi.org/10.1177/1098214005283748.

Thomas, G., 2011. A Typology for the Case Study in Social Science Following a Review of Definition, Discourse, and Structure. Qualitative Inquiry, 17(6), pp.511–521. https://doi.org/10.1177/1077800411409884.

Titlestad, M., 2013. Searching for the Sugar-coated Man. Safundi, 14(4), pp.466–470. https://doi.org/10.1080/17533171.2013.841061.

Tong, L., Lindeman, R.W. and Regenbrecht, H., 2021. Viewer's Role and Viewer Interaction in Cinematic Virtual Reality. Computers, 10(5), p.66. https://doi.org/10.3390/computers10050066.

Tricart, C., 2017. Virtual Reality Filmmaking: Techniques & Best Practices for VR Filmmakers. [online] Taylor & Francis. Available at: <https://books.google.ie/books?id=FN5CDwAAQBAJ>.

Tseng, C., 2013. Cohesion in Film: Tracking Film Elements. Springer.

Tseng, C. and Bateman, J.A., 2012. Multimodal narrative construction in Christopher Nolan's Memento a description of analytic method. Visual Communication, 11(1), pp.91–119. https://doi.org/10.1177/1470357211424691.

Tseng, C.-I., Laubrock, J. and Bateman, J.A., 2021. The impact of multimodal cohesion on attention and interpretation in film. Discourse, Context & Media, 44, p.100544. https://doi.org/10.1016/j.dcm.2021.100544.

UNHCR, 1999. Un Voyage Pas Comme Les Autres Sur Les Chemins de l'Exil" (A Voyage Unlike Any Other on the Road to Exile). [Immersive Theatre].

United Nations, 2018. UNVR United Nations Virtual Reality. [online] Available at: <http://unvr.sdgactioncampaign.org/> [Accessed 5 August 2018].

Unsworth, L., 2006. Towards a Metalanguage for Multiliteracies Education: Describing the Meaning-Making Resources of Language-Image Interaction. English Teaching: Practice and Critique, 5(1), pp.55–76.

Ure, J., 1971. Lexical density and register differentiation. Applications of linguistics, 23(7), pp.443–452.

Uricchio, W., Ding, S., Wolozin, S. and Boyacioglu, B., 2016. Virtually There.

Van Leeuwen, T. and Jewitt, C., 2001. The Handbook of Visual Analysis. [online] SAGE Publications. Available at: <https://books.google.ie/books?id=vSFlLmCR26oC>.

Ventola, E. and Guijarro, A.J.M., 2009. The World Told and the World Shown: Multisemiotic Issues. Springer.

Vertov, D., 1984. Kino-Eye: The Writings of Dziga Vertov. [online] University of California Press. Available at: <https://books.google.ie/books?id=U60wDwAAQBAJ>.

Vice News, 2015. Chris Milk, Spike Jonze, and VICE News Bring the First-Ever Virtual Reality Newscast to Sundance. [online] Available at: <https://news.vice.com/article/chris-milk-spike-jonze-and-vice-news-bring-the-first-ever-virtual-reality-newscast-to-sundance> [Accessed 27 July 2017].

Walker, L., 2016. The Vodou Healer. [online] WITHIN. Available at: <http://with.in/watch/the-vodou-healer/>.

Waugh, T., 2011. The Right to Play Oneself: Looking Back on Documentary Film. Visible evidence. [online] University of Minnesota Press. Available at: <https://books.google.ie/books?id=y5j8D8GBJywC>.

White, G., 2012. On immersive theatre. Theatre Research International, 37(3), pp.221–235. https://doi.org/10.1017/S0307883312000880.

Wingstedt, J., Brändström, S. and Berg, J., 2010. Narrative Music, Visuals and Meaning in Film. Visual Communication, 9(2), pp.193–210. https://doi.org/10.1177/1470357210369886.

Winston, Brian., 1995. Claiming the real: the Griersonian documentary and its legitimations. London: British Film Institute.

WITHIN, 2019. Within. [online] Within. Available at: <https://www.with.in> [Accessed 24 January 2020].

Yin, R.K., 2009. Case Study Research: Design and Methods. Applied Social Research Methods. [online] SAGE Publications. Available at: <https://books.google.ie/books?id=FzawIAdilHkC>.

Zhang, Y. and Weber, I., 2021. Adapting, modifying and applying cinematography and editing concepts and techniques to cinematic virtual reality film production. Media International Australia, p.21.

Appendices

Appendix A: Pilot Analysis of The Vodou Healer

Metafunctional Analysis of Phase 1 Phase 1.1 Ideational

The ideational metafunction denotes a state of affairs within the field of existence in which the text intersects. The image-auditory combine in this instance through the conjunctive image-text relations of *extension* (van Leeuwen, 1991) in that the visual-spatial modes and the auditory (spoken) mode are not the same and thus *added* but with no explicit stated purpose (ibid.).

The voice-over presents a transitive actor-goal dynamic within a *participant-process-circumstances* frame. The earthquake is a presented through the active verb 'made' as an actor enacting a *causal* process on the speaker ("making me live, to see") in circumstances of unforeseen disruption to a notional status quo ("things that I never thought I would see"). Temporally, the process exists is in the past tense where ideational meaning is that of a living survivor of a relatively recent earthquake that has brought significant as-yet undefined (for us) upheaval in her life.

Simultaneously, the visual-spatial modes present a homogenous field of activity with a high level of naturalness, exhibited by the many mini-transactions (both literally and figuratively) indicative of daily market existence. The gaze, movement and proxemics are suggestive of the scene carrying on un-impinged by the presence of the camera. Ideationally, the naturalness and non-hierarchical series of movements suggests that 'this is the status-quo, in this place that you are standing'.

The auditory and visual modes combine to overlay the space with an extended unseen ideational meaning connecting the past (the earthquake) to our embodied position in the perceived present. Thus, the status of what we are seeing becomes indeterminate (that it is 'normal' but only by our unawareness of an unknown pre-earthquake standard).

Phase 1.1 Interpersonal

The spectator is positioned statically and at a natural standing height and can thus construed as a human participant of the scene. In the lack of visual address in the participants (gaze) we are not privileged in the space and the fluid movements by those around the camera at varying close (social and personal distances) enforces this status. This subphase contains no observable visual reference to the speaker and therefore certain embodied relations are not

at varying close (social and personal distances) enforces this status. This subphase contains no observable visual reference to the speaker and therefore certain embodied relations are not available for analysis.

The interpersonal dimension is approached instead using the above observations on an implied spectator inferred through the auditory track. Much can be derived from the linguistic construction of the voice-over in and of itself and for how music adds paralinguistic signification to the speaker. Using Goffman's participant framework, the spectator can be constructed by the speaker's utterance The speaker's verbalized position is established via their linguistic choices in a number of ways. Firstly, using the SFL *mood* and *modality* markers (the 'stance' taken by the speaker) we see that in this instance she is using the *indicative; declarative*. This modality is that of *presenting information* from the subjective, experiential standpoint. The information provided can be further identified through its use of the self-referential past-tense ("made *me…I* never thought *I*") as *autobiographical*. In doing so she is momentarily presenting herself as *narrator* through a form of biographical account and in doing so the spectator is given the status of 'story listener' (ibid. p. 22).

The somewhat paradoxical clause construct relating to how the earthquake 'made her live' characterizes her autobiography as both victim and survivor simultaneously. As she speaks the words, the music track switches to an elongated single *tonal cluster* a chord structure with no root-note or pitch hierarchy (Rick, 2016) that adds sonic indeterminacy and dissonance to the auditory track, supporting the speaker's identity as having been thrown into a state of flux.

Phase 1.1 Textual

Textual cohesion is established in the auditory track on two levels; across the music-voice components and then within the voice-over itself. Temporally, the music track 'book-ends' the subphase in question thus framing the voice-over. The music begins with a slow rising two-chord progression, the second a melancholic minor chord ringing out to the point where the speaker commences. Following the voice over the chords repeat with the addition of an additional chord that adds new auditory information acting as a bridge between subphases 1 and the following subphase, coinciding with the visual cross-dissolve effect that forms an 'explicit conjunction' (van Leeuwen, 1991) to the next visible location.

In the voice-over, cohesion is established syntactically with the earthquake established as the topical theme and elaborated upon across the clause structure (rheme). Information thus builds across the clause structure in a wave-like fashion from the small waves of information ('the

earthquake made me live' ... 'to see things' ... 'I never thought I would see'). Martin and Rose (2003) posit that the textual function must also to establish and position participants in the text (and track them thereafter). Here, an as yet unnamed speaker is identified ("me", "I") as earthquake survivor and vehicle for us to know the effects of the disaster.

In the visual-spatial modes the spatial plane extends from our embodied position outwards, itself a wave of new information start with the spectator and ending at the distant edges of the market (and in paces, the distant hilltops). The kinesic mode connects the visual-spatial plane also as the movement of participants leads the spectator's gaze around to the scene. When the auditory and visual-spatial are considered together, a cohesive tie of verbal rhythm in the voice-over, through the repetition of the word 'see' creates cohesion between the embodied gaze of the spectator to the heard voice. Spatially, our fixed available perspective on the space, added to the open ended clause "things I never thought I would see" imposes a reality beyond the edges of the space that we are not privy to, thus setting up a cohesive tie with the following subphase.

Bridging subphases 1.1 and 1.2

Here I will present a single analysis of both subphases 1.2 and 2.1. A textual transition occurs in the piece in which the trajectory of meaning is significantly altered *within* a shot. As such space-time is continuous and undisrupted across the subphases, allowing for continuous description of events around and through the transitional moment. Such approaches are necessary when phases are identified in the 'discourse stratum' which does not always line up neatly with the formal units of video based media. I propose that the first phase ends with the mention of her husbands passing and the conjunctive 'but' begins a new phase (2.1) identifiable across all metafunctions.

Phase 1.2 / 2.1 Ideational

As with the previous subphase, the voice-over is located grammatically in the past tense. The verbal clauses depict a single participant (the speaker's deceased husband) in the temporal and causal circumstances of the earthquake. The existential process of dying is not explicitly attributed to the earthquake as causation, but the conjunction "right after" joins the two clausal components tying the happenings together, if in an open fashion. The following conjunctive "but" is significant in that it shifts matters explicitly into the future tense. The clause thereafter contains the mental process of 'knowing' that a ceremony is imminent and that her deceased husband will be present in a spiritual sense.

3

The visual-spatial modes in part *illustrate* what is said and in doing so *confirms* the verbal address (van Leeuwen, 1991, p.92). As the first clause "Right after the earthquake, my husband died" is voiced, Katy kneels at what appears to be a gravestone or monument. In terms of her gaze, she looks down in an engaged manner involved in a transitive reaction to the grave-stone object (Tan in Hyland & Paltridge, 2011, p. 129). The following clause "But I know his spirit will be at the ceremony" is illustrated by her switching her gaze to an upward vector in a more disengaged fashion implying what Tan describes as 'self-involvement' indicative of primarily internally directed mental process (ibid.). This internalized process is reinforced in the kinesic mode by her rhythmic swaying.

The ideational functioning here in phase 1.1 is to add detail to the earthquake and its repercussions.

Phase 1.2 / 2.1 Interpersonal

The speaker is introduced in this phase visually as the sole visual participant, and also formally identified via a text graphic. The voice remains in the *indicative-declarative* form regarding the earthquake, the death of her husband and his imminent spiritual presence at a ceremony.

Our proxemic relations to Katy are on the boundaries of the personal-social. The camera is positioned low relative to the previous shot indicating a conjunction between Katy and spectator of 'seated together'. The initial albeit slight movement of Katy 'settling' into the kneeling position can be construed as her joining the preexisting spectator at the grave. Katy's engaged gaze to the gravestone sets up a visual relation with 'husband' and through that visual conjunction the spectator is given an ambiguous participation status of either *bystander* or *eavesdropper*. Combined with the ever-present voice however the status can be considered in Goffman's terms to be *ratified* (Goffman 1979) to some real degree.

The voice track remains in the *indicative mood*, again providing historic information via subjective statements. Matters shift after the 'but' conjunction however when Katy moves instead to propose with a medium to high level of certainty (modal adjunct "I know" (Halliday and Mathiessen in Aijmer, 2016)) that her husband will be at the ceremony in spiritual form. This addition of her own mental process of *knowing* establishes Katy as 'expert in these matters'. The use of the definite article ('the ceremony') assumes a familiarity between the spectator and matters at hand. Goffman's frameworks would class this juncture as a change of *footing* enacted by Katy (Goffman 1979), in which the situation is reframed from historical story-telling to a mode of speech that involves the spectator directly in matters at hand; bringing the already initiated

spectator 'up to speed' so to speak.

Phase 1.2 / 2.1 Textual

The verbal track connects backwards to the previous subphase by initiating the clause with the pre-established earthquake (a 'recoverable theme' (Martin, 2002)), adding the information of her husband's death and relating future spiritual and ceremonial happenings to historical context through the relational conjunction 'but'. This relational conjunction acts as a nexus for the larger phasal and textual formations that elaborate the ceremony, and Katy as transformational agents in historical circumstances. The husband thus becomes the *topical theme* ('but his') that is elaborated through the proposed circumstances of his presence at the ceremony.

As mentioned previously, the music track conjoins this and the previous subphase. The added chord rings out across the two locations, the harmonic chord elevated from the previous with this one through the introduction of chord progressions that play out across the transition between shots and ring out. The tonal cluster employed in the first subphase is reiterated here connecting the voice-over passages tonally and thematically.

Discussions from the Analysis

A simplistic summation of the analysis in terms of its interrelated metafunctions can be expressed as follows.

The earthquake has altered the reality of Haitian life; Katy, our Haitian 'guide' is a victim of tragic events but maintains a degree of hope and thus agency over her circumstances. As standing in a post earthquake Haiti, we as spectator cannot see what Katy saw but is now enjoined to become part of the ceremony.

Such a summation leaves much to be established across the broader text but is interesting in the manner in which this state of affairs is realized in the brief account of the first subphases.

Broadly, the relations of the auditory and visual-spatial modes present a possible re-ordering of the image-text hierarchy as put forward earlier (Van leeuwen's *extension*, whereby the image tends to extend the voice-over in direct-address). The subphase 1.1 (row 2) in which we are positioned in the quotidian of market life, engages all spatial modes to a high degree (and with it our cognition) and gives the visual-spatial a high level of *salience*. The voice serves to extend or 'overlay' meaning on the *place* in which we are 'standing'. Hence, rather than the image further signifying what is said, the directionality of the relations can be considered to be reversed.

5

The earthquake and the speaker's biographical account (as spoken in the voice-over) prompt us to take what is given before us but imbue in the perceived space a social or experiential reality beyond the 'facts' before us.

Also of note here is in the textual conjunction of word in auditory mode and the spectator's gaze. The repetition of 'see', creates cohesion that ties the visual-spatial reality (what we see) to the experiential reality of the speaker ('things I never thought I would see').

With regard to the role of the spectator, there is in one sense an embodied positioning of the camera from which eyes and ears and thus a spectator-body can be inferred. The proxemic- spatial mode shows this effect in rows 2 and 3 where the camera height and interactions with the space suggest a standing and sitting position, respectively. Beyond this Goffman provides interesting frames for considering the manner in which the spectator is further cast into more *psychological* roles. The change of footing employed by Katy in the phase transition for example (row 3: 'but I know…') sees a shift in spectatorial role from the more passive 'story listener' spectator role to the more *included* role of ceremonial attendee. This is reflected in the simple switch from past to present and future coincide in this moment that is purely linguistic. This provides potential for a more holistic account of hat it means to be 'embodied' interpersonally in a scene.

The interpersonal dimension regarding Katy's embodiment raises interesting questions also. At various points she is either heard as voice-over (and *not* seen), seen *and* heard (but heard only as voice over) and in later phases she is seen and heard as both her synced voice and with voice-over. While this is not uncommon in framed productions, the level of *presence* assumed between a seated spectator and Katy in the CVR-space must be considered to change the connection between us, Katy and her dislocated voice. As such it could provide interesting avenues of inquiry relating to an *embodied person* with *disembodied voice* dichotomy suggestive of hearing 'private thoughts' belonging to CVR participants that is different to our typical reception of direct-address. This has further ramifications regarding Goffman's notions of ratified participants; those with designated 'access' to aspects of a situation.

Connected with this, ambiguity arises also in *our* access to Katy's voice that we must assume is not available to the other inhabitants in the scene. Goffman phrases his participation framework as premised on a situation whereby a 'social situation [is] the full physical arena in which persons present are in sight and sound of each other' (Goffman, 1979, p.10). 'Sight and

6

sound of each other' in CVR is of course problematic in this regard.

Concluding Remarks on the Pilot Study

As an experimental foray into multimodal discourse analysis on a CVR artefact, this study has proven useful in highlighting a number of practical and theoretical factors that will need to be addressed. Firstly the practical aspects regarding transcription of CVR for analysis requires further experimentation. Criticisms of subjectivity potentially undermine matters with regard to the visual references used in column 3. Possible solutions may involve the inclusion of graphic schematic representations of a location's layout, yet these also lack a degree of empirical rigor. Alternatively, additional 'extracted' frames from more perspectives may serve to provide more descriptive resolution. Considering issues such as this highlights the potential scale for analyzing multiple CVR films. Having performed this study in a very limited form, I am better situated in finding the required balance in terms of the amount of films, duration of films and the granularity of analysis.

Added to this, other transcriptional elements will require more consideration with the spatial-proxemic mode as a case in point. Proxemics is indeed an important factor in analyzing immersive social situations. Foregrounding it perhaps hinders the overall spatial factors that will serve the study more efficiently. 'Layout' might be more apt in describing the spatial mode that can include proxemic as one of its metrics.

A methodological issue encountered related to foregrounding of language analysis (the voice-over). This is partly owing to the SFL's maturity as a technique for analysis and rich set of resources available to the analyst, but runs contrary to MMDA principles that eschew *language-centred* approaches with all else considered *paralinguistic*. In performing analysis on the spatial aspects of the scenes, the tendency was to fall back on a quasi-linguistic interpretations. I consider this an issue of proportionality that can be addressed in tipping the balance towards theories of space (which inevitably involve the body) such as those put forward by Peirce, Ponty and Le Febvre.

Appendix B: Filtering Process from Mediography

Note		Location	Producers origin	3rd party orgs	Theme: Mediography (my additional themes)	Contextual information	Subject speaks	inter-subject dialogue	Talks to camera	Reflexive/Transparent
y	7 Stories for 7 Years: Life After Syria	Syria		World Vision	Refugees & Syrian Civil War	Narrator (Anglo)	y/n	n		R
	Amazigh				Climate change, Migration, Nomads, Peoples & Cultures & Sahara	Narrator	n	y (background)		T
y	A Family's Return to Mosul	Iraq		NYT	War & Conflict	Interstitials	y	n		T
y	Aftershock: Nepal's Untold Water Story	Nepal		Water Aid Amor de	Earthquake, Environment, Natural disaster & Nepal	Narrator	y (translated)	y (untranslated)	y	R y
y	Abuela (A Grandmother's Love)	Guatemala		Brightlight Foundation	Guatemala, Poverty & Solar Energy		(subtitled)			T
y	Behind the Fence	Myanmar		The Nexus Fund Bridgewater Foundation UN Population Fund (mentioned not credited) UN, UNICEF (thanked)	Burma, Ethnic cleansing, Human rights, Myanmar & Rohingya	Narrator	y	y		T
y	Born Into Exile	Jordan			Migration, Refugees & Syrian Civil War	Narrator	y (translated)	y (translated)		R/T
y	Clouds over Sidra	Jordan			Migration, Refugees & Syrian Civil War	Narrator	n		y	R/T
y	Cut-Off VR (non-naturalistic)									
n	Ch'aak' S'aagi (trailer only)									
y	Congo VR: A Troubled Past	Congo		BBC	Africa, Congo, Congo River & Journalism	Narrator	y (professional/subject transl.)		y	R
	Congo VR: War and Disease	Congo		BBC	Africa, Congo, Congo River & Journalism	Narrator	y (professional/subject transl.)	y	y	R
	Congo VR: Great Riches	Congo		BBC	Africa, Congo, Congo River & Journalism	Narrator	y (professional/subject transl.)	y	y	R Crossing
y	the Sky: Trek to School through the Himalayas	Nepal		(BBC production)	Nepal	Narrator	y		y	R
y	Duppy Gun Riddims									

	Title	Country	Organization	Topic	Narrator				
y	Finding Hope in the Vigils of Paris: A Virtual Reality Film y Finding Haka								
	From Waste to Taste	Brazil	Brazil & Qatar		Narrator	y (translated)	n		
y	Happyland360	Phillipines	CREST	Philippines, Poverty & Slums		y			R
y	I Am Rohingya	Bangladesh		Ethnic cleansing, Migration, Myanmar & Rohingya	Narrator	y (translated)			n Jisr al-
y	Shoughour, a devastated Syrian city	Syria		Syrian Civil War & War & Conflict	Narrator	y (translated)			R UNHCR
y	Life in the Time of Refuge Monster | Oculus VR for Good Creators Lab	Lebanon/Finland	The Humanitarian Cooperative	Migration (Syrian conflict)	Narrator (acted)	n	y (translated/untranslated)		R Meeting a
y	My Mother's Wing	Gaza	UN	Gaza & War & Conflict	Narrator	y (translated/untranslated)	y		R/T
y	Mamie's Dream	Sierra Leone	Plan International	Crime & Justice & FGM Equality, Government & Politics, Law, Same Sex Marriage & United States	Narrated (Mamie?)	y	n		T
y	Marriage Equality VR	USA							
y	My Africa	Kenya	Conservation International	Africa, Conservation & Kenya	Narrator	n	n		- y
y	Nasra	Somalia		Female empowerment & Somalia	Narrator	n (untranslated)	n		R
y	Notes to My Father	India?		Grief, Human trafficking & Sex Trade	Narrator	y	n		T
y	On the Brink of Famine	S. Sudan	Frontline, Medicin Sans Frontiere, Brown	Famine, Sudan & War & Conflict	Mixed	y	n		R
y	Oil in our Creeks								
y	Our Home, Our People	Fiji	EU/ACP World Bank (climate)	Climate change & Fiji		y	n		y
	Paramount Ranch's Western Town destroyed by the Woolsey Fire			Natural disaster, Wild Fires & Woolsey Fire 2018	None	n/a	n/a	n/a	T
	Paris' Queen Bee	France		Beekeeping & Bees	Narrator	y (translated)	n	y	y U
y	Pencils of Promise	Ghana	S A RYOT	Africa, Awareness, Education & Ghana	Narrator	n	n		R
y	Portrait of a Caregiver NYT								

	Title	Location	Country	Producer/Organization	Topic	Narration	Language				Rating
y	Refugees	Lesbos	Netherlands		Migration, Refugees, Syrian Civil War & War & Conflict	Narrator	y (Afghan: speaks English)	n			R
n	Rise Above										
y	Remembering D-Day										
y y	The People's House	　									
	The Contenders	NYT platform									
	The Okavango Experience: Episode 1	Africa, Botswana, Nature & Okavango Delta	USA	Black Dot Films VR, Google, National Geographic & Q Department		Narrator	n/a				
	The Okavango Experience: Episode 1	Africa, Botswana, Nature & Okavango Delta	USA	Black Dot Films VR, Google, National Geographic & Q Department		Narrator No	n/a				
	The Oak Forest Mobile Estates neighborhood after the Woolsey Fire				Natural disaster, Wild Fires & Woolsey Fire 2018	voice	n/a y (translated)	n/a	n/a y		T R
	Unidentified Sliding Object				Hovercrafts, Racing & Sport	Narrator					
y y y	Vice News VR: Millions March NYC Warwick Gold - Australian Rodeo	Australia	Australia		Extreme Sports & Rodeo When	Narrator		y n			T T
y	Land is Lost, Do We Eat Coal?	India	Amnesty International India	Coal Mining, Displacement, India & Korba London 7/7 Attacks &	Narrator	y (subtitled) y	n			R
	Witness 360 - 7/7	UK		Terrorism		Narrator					
y	Waves of Grace	Liberia	US UN		Ebola & Health	Narrator	y (translated)	y		T	
y	Yeh Ballet	India	India		Ballet & Performance	Interview w-cutaways	y (interviews)	y (interviews)		R	
y	Yemen's Skies of Terror	Yemen	Al Jazeera		War & Conflict & Yemeni Civil War	Narrator	y (translated) y/n			R	
y	Iranian Kurdish Female Fighters	no video - seen it in conve rsatio n									
	No Small Talk Collisions	Jaunt platform only not in 3D									
y	Home: Aamir										
y	Nepal Quake Project										
y											

					y (translated)		
y	The Vodou Healer	Vaseline	Magic, Paranormal & Voodoo	Narrator		y	R/T

syria	Nobel's Nightmare						
sea n - RYOT	Palmyra Atoll 360						
find	Purely Peru	USA					
y	Zika Virus: Inside the Epidemic's Center in Recife	Brazil		Brazil, Journalism & Zika Virus	y (translated)	??	R
	We Shall Have Peace						

Appendix C: Examples of Transcribed Shots

Appendix C.1 Transcription of Congo VR Shot 17

1	**VISUAL FIELD**
2	
3	**Gloss:**
4	[dissolve:s]
5	Village surrounded by tall trees, comprising thatched huts. Row of men moving rhythmically in tandem, some holding machete style knives and pipe-like objects; sparse onlooker stand by. 2 production personnel hold equipment; camera operator films men. Green object of woven leaves on ground at below, close to camera stand on ground.
6	[dissolve:s]
7	**Kinesis**
8	{CLUSTER #1 TRIBESMEN {11} moving rhythmically in synchronicity with arms swaying {m}; DRUMMER strikes a beat with knife on large {{bamboo-like}} object {m}}
9	BBC WOMAN walks to opposite side of TRIBESMEN looking at phone {s}; steps closer to A.L
10	A.L walks after BBC WOMAN and joins her
11	SITTING MAN walks to plastic chairs; sits; swings leg over chair arm {m}
12	WOMAN WTH BABY adjusts baby on her back {f}
13	STANDING MAN shifts weight onto other leg and puts hand on hip
14	General movement of CROWD at opposite side of road as they position themselves to view unfolding scene
15	**Spatial/Proxemic**
16	[CAM P] on ground, on front of row of CLUSTER #1 TRIBESMEN
17	[CAM H] med
18	[activated hor.] enclosed {tight} on one side by TRIBESMEN and buildings {near} giving on to tall trees; open {far} to roadway but obscured on by trees and other people
19	[activated vert.] open to clouded sky
20	[CAM <> TRIBESMEN] personal; front > social; front-side
21	[CAM <> A.L] social; side >> personal/social; side >> social/public; side
22	[CAM <> BBC CREW] social/public; mixed
23	CAM <> CROWD {{across road}}] public; front
24	[CAM <> MAN IN KHAKI] social; front
25	[> CLUSTER #1 TRIBESMEN <] intimate; side {{lined up}}

26	**Gaze**
27	CLUSTER #1 TRIBESMEN (11): Downward and toward each other (mixed – disengaged)
28	CAMERA MAN: toward TRIBESMEN (engaged)
29	BBC WOMAN: toward TRIBESMEN (engaged); down toward phone (engaged)
30	A.L: toward TRIBESMEN (engaged)
31	CROWD: towards scenario (engaged
32	STANDING MAN: towards tribesmen (engaged)
33	**Diegetic Sound:**
34	[✿] (...cont) rhythmic chanting; beating of single percussive instrument (vol: h)
35	**Music:**
36	[✿] (...cont) rhythmic chanting; beating of single percussive instrument (vol: h) (17_VILLAGE TRIBE, Pos. 34)
37	**LEXICOGRAMMAR**
38	47.37%
39	**This is (Buyanger) village. They're performing a good luck ritual before a forest hunt ...and we're going with them**
40	**Prosody**
41	This is (Buyanger) village↓(.5) they're performing a good luck ritual↓(.) before a forest hunt (.) and >we're going with them<
42	.1. This is /Buyanger /village.
43	.1+ They're /performing// a /good /luck ritual // before a forest hunt
44	-1. and we're going /with them
45	**Ideational**
46	**Participants:**
47	1) token: (this); Value: Buyanger Village
48	2) actor: (they); range: (a good luck ritual)
49	3) actor: we
50	**Processes:**
51	1) relational/identifier: (is)
52	2) material: (performing)
53	3) material: (going)
54	**Circumstance:**
55	2) Location/time: (before a forest hunt)
56	3) accompaniment (with them)
57	**Logico-semantic**
58	(1) This is (Buyanger) village ✕ (.5) ||| (+2) they're performing a good luck ritual ✕ (.) before a forest hunt (.) ||| (+3) and >we're going with them<
59	**Interpersonal**
60	MOOD
61	1) Unmodulated declarative: this is (Buyanger village
62	2) Unmodulated declarative: they're (performing)
63	3) Unmodulated declarative: we're (going with them)

64	**Textual**
65	{unmarked topical: this} > is
66	{unmarked topical: they} > 're
67	{unmarked topical: we} > 're
68	Identities:
69	Buyanger village
70	Pygmies {they} [superstitious hunters]
71	The team and I + spectator

3

Appendix C.2 Transcription of Mamie's Dream Shot 13

1 **VISUAL FIELD**

2

3

…start probability frame

4 **RP**

5

…Reading path

6

End probability frame

7 **Gloss:**
8 [/dissolve:s]
9 Mamie, smilingly walks across a busy outdoor social area towards, while conversing with a woman in the centre of the area cooking in a pot over a fire. The area is busy with a crowd of people; some children observing Mamie, while most engage within smaller clusters.
10 A group of youths line up in a circular formation and walk clockwise around
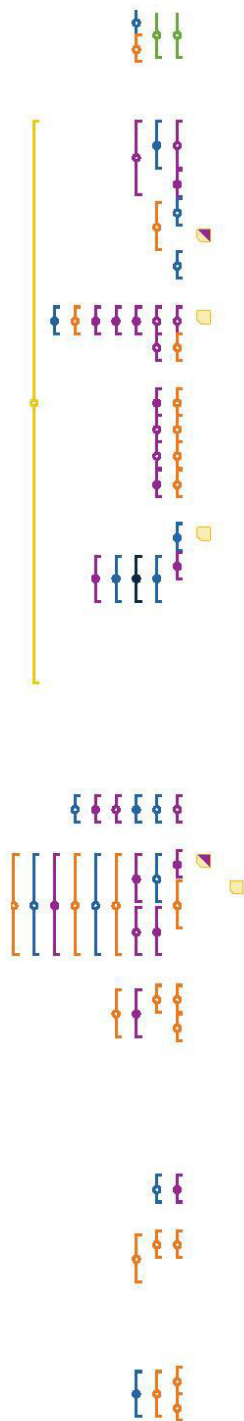
the perimeter of their self-made circle.

11 [Dissolve: m

12 **Spatial/Proemic:**
13 [CAM P: moving] {(moving along ground, approx. 20 feet)} along centre of open area
14 [CAM H] medium
15 [activated hor.] rectangular-open enclosed by trees and buildings {one side mostly}; open at either end leading along to distant buildings
16 [activated vert.] enclosed overhead by light tree cover

17 [CAM <> MAMIE] social/public; front >>> public; front
18 [CAM <> COOKING WOMAN] public; side >>> social/public; front >>> public; side
19 [CAM <> CLUSTER #1] public; front >>>
20 [CAM <> CLUSTER #2] >>> personal/social; mixed >>>
21 [CAM CLUSTER #3] public; mixed
22 [CAM CLUSTER #4] >>> public

23 [MAMIE: side; away (moving) <> CLUSTER #1: toward] public >>
24 [MAMIE: toward-side; away (moving) <> COOKING WOMAN: toward -side] >> personal >>
25 [MAMIE: toward -side <> CLUSTER #2: front-side/front] >> personal/social >>
26 [MAMIE: toward <> CLUSTER #4: mixed] >> public

27 **Kinesis:**
28 [
29 [CAM Movement; m; cinematic: tracking Mamie with slight shake]

30 {MAMIE walks through yard; ^ gesticulates towards CLUSTER # 1 people seated at edge of yard; pauses at gesticulating conversationally with COOKING WOMAN; continues walking, waves in direction of seated woman} {m}

31 {COOKING WOMAN adjusts cooking pot; picks up a large ladle/spoon; ^ looks up at MAMIE; lifts lid of cooking pot and stirs contents of pot} {m}

32 CLUSTER #2 MOSTLY FEMALES
33 {General movement in standing/seated group: WOMAN lifts lid on plastic container; WOMAN rearranges infant on lap} {m}

34 CLUSTER # 3 MALES
35   [
36   MAN #1 W/ FOOTBALL kicks ball - solos {('keepie-uppie)}

37   ]

38   BOY WITH WHEEL walks from distance and rolls his wheel to MAMIES position {m}

...Framing

...transacti

...Kinesis

...Proce

...Circur

...Of

...Ca

...Gaze

...Identi

...Framing

...Central > m

...Reactional

...demand,

...Reaction

...manner

...Diegetic sound
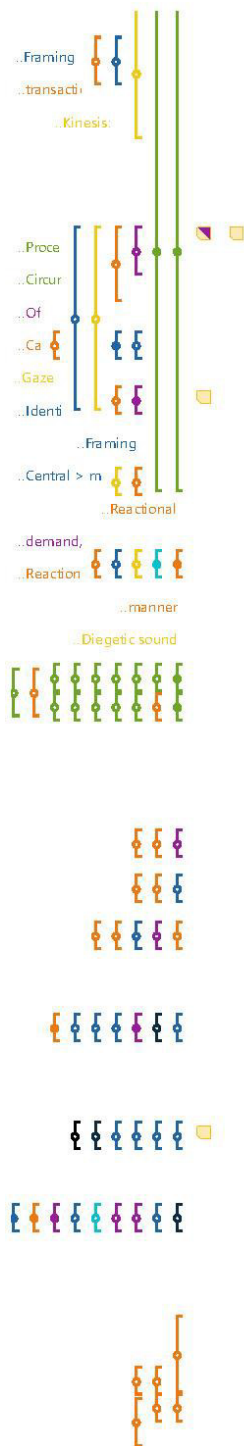
| 39 | CLUSTER #4 CHILDREN |
| 40 | CHILDREN march in a |

circular formation {m}

41 General movement indicative of social event with mixed ages {mixed}

42 Gaze:

43 MAMIE: ^ to CLUSTER #1 people seated at edge of yard {engaged}; to COOKING WOMAN {engaged}; to CLUSTER #2 {engaged}

44 Cooking woman ^ M

45 CLUSTER #2: mostly within cluster at pre-existing interactions {mixed}

46 CLUSTER #3: downward and inward at ball activity {engaged}

47 CLUSTER #4: inward at activity or walking a circle {engaged}

48 CLUSTER #5: at camera/Mamie

49 Diegetic Sound:

50 [⚙] lively voices (vol:m)

51 Music:

52 Interlocking hand-claps added to instrucmentaiton

53 LEXICOGRAMMAR

54 People used to mock me and call me names. They said nobody will ever respect me. But now I am a community leader and a role model

55 Density: 51.85%

56 Prosodic

57 People used to mock me↑ {.} and call me #names#↓

58 They said nobody will <ever respect me>

59 [vo] But now {.} I am a community leader ↑{.} and a role mo:del ↓

60 Intonation:

61 .2+ People used to /mock me // .1. and /call me /names

62 Low PT to sharp rise, neutral, continuative incomplete // neutral fall statement

63 .1. ^They said /no/body will /e/ver /respect /me

64 Slow fall from 'no' problematic as tonic is unclear and falls like para

65 .1- But /now // I /am /a community /leader // .1. and /a /role /model

66 Statement // step down PT to sharp rise cont. incomplete // statement – lands on 'wrong' tonic

67 Ideational:

68 Participants:

69 1} sayer/behaver: people; me: carrier; attribute: names

70 2} sayer: they; verbiage: nobody will ever respect me

71 3} carrier: I; attribute: a community leader/role model

72 Processes:
73 1} verbal: mock me/call me names
74 2} verbal: said
75 3} relational: {I} am

76 Circumstances
77 4} time: now

78 **Logico-semantic:**

79 1 People used to mock me // +2and call me names ||| They said nobody will ever respect me ||| But now I am a community leader and a role model

80 **Interpersonal:**
81 Mood: Declaratives:
82 1} People used to mock me {and...}
83 2} They said {nobody will...}
84 3} I am {a community leader}

85 **Textual:**
86 1} topical: People > used to mock me and call me names
87 2} topical: they > said nobody...
88 txt: but) {top: now > I am a community leader

Left margin labels (diagram):
..Verbal
..Verbal
..Relational
..time
..Voiceover/music-track
..mu
Inte
..re
..mood
..residue
..Declarative
..residue
..mood
..topical
..topical

Appendix C.3 MAXQDA Interface

A. Internal documents housing transcripts of individual shots

B. Code systems (left) and coded segments of data

Appendix D: Annotation Key

Annotation		Modifier/examples	
(())	Analyst in-text comments		
CLUSTER	Groups of participants		
PARTICIPANT NAME	All human actors denoted in all caps	SEATED WOMAN	
Mixed	Properties in flux		
/	Potentially either-or properties on either side	**Social/public** (may be either, not possible to discern)	
Visual Content			
[GFX titles]	Graphics used at beginning and end of film		
[GFX]	Additional graphic information used to assist the narrative		
[Dissolve]	Inter-shot dissolve with next shot	[Dissolve: s, m, f]	Dissolve: tempo (slow, medium, fast)
[\Black]	Fade down to black	[\Black: m]	Fade to black: tempo (slow, medium, fast)
[Black/]	Fade up from black	[Black/: m]	Fade up from black: tempo (slow, medium, fast)
Audio track			
[VO]	Denotes voice-over		
[ ] (wd+R)	Denotes diegetic location sound	rustling leaves; **shouting**	Bold indicates sound' salience
		undefined	Sound not identifiable
[♫]	Denotes music track (excluding diegetic, heard music)		

vol:	Music volume (low, medium, full)		
tempo: s, m, f	Music/voice's tempo (slow, medium or fast)		
Voice/prosody			
(.5)	Denotes gaps in seconds, pauses in spoken tracts.		
=	Connects speech units run-on (no gaps)		
[]	Overlapping speech between participants		
underlined	Volume increased		
CAPITALIZED	Denotes volume emphasis on specific words		
::	Used within words to indicate elongated vowel segments		
↑(wd3+h)	Elevated pitch		
↓(wd3+i)	Lowered pitch		
//	At end of spoken tract to indicate a closure		
<>	Tempo of words between angles decreases; words stretched out		
><	Tempo of words between angles increases; words compressed		
°°	Words between degree symbols are quieter		
##	Words between hash symbols are softened and 'creaky'		
()	Speech unidentifiable		
Gaze			

Focus	The object (where applicable) of the gaze is noted where no object, direction is used		
(engaged)	Focused gaze		
(disengaged)	Unfocused gaze		
^	Is reacting to something/someone		
(ind.)	Indeterminate. Not discernible where or how they are looking		
Spatial-Proxemic	Distance between camera and participants Distances between participants Spatial descriptions of environment		
[CAM P] [CAM P:moving]	Camera position (horizontal plane) Camera repositions in-shot		
[CAM H]	Camera height	Sub: lower than ground level x-l: ground-level L: 1' – 3' m-l: 3' – 5' m: 5' – 6' m-h: 6' – 12' h: 12' – 20' x-h: 20' > aerial/building roofs etc.	
[Activated hor.]	Activated space on the horizontal plane. 'open space' relative to the camera	180° Enclosed: Tight (< 2m) Enclosed: Near (3m – 20m) Open: Far (21m – 100m)	

		Open: Horizon (undetermined/far off)	
[activated vert.]			
<>	Space between salient participants/objects including orientation (towards, side, away)		
[CAM <> participant/object: f, s, a]	Distance from camera to participants including their orientation to camera/each other (front, front/side, side, away/side, away	(**PERSON#1**: personal/social; front-side)	Person(s) related by distance from camera position
		(PERSON#1: **personal/social**; front-side)	proxemic value: may be border distance between two proxemic values
		(**[PERSON#1+PERSON#2]**: personal/social; front-side)	camera distance relative to more than one person
[> CLUSTER <]	General distances in field of participants		
>>>	camera moves during the shot		
Angle	Angle of persons relative to camera/each other/objects		
Kinesic Action	N.L descriptions of: participant movements; movement of environmental components; camera movement (not noted if static)	s, m, f	speed of movement from slow to fast
^	Conjoined, reactive movements and actions *between* participants. Also where movement implies cognisance of camera		
;	Semi-colons connect discrete sequential movements		

()	compound movements denoted within parenthesis		
Bold	Salient movements in bold		
[ ]	simultaneous movements		
Indented text	Passages indented from left to right imply sequential timing between complex kinetic sequences and groups of simultaneous movements		
[CAM Movement]	Implied camera movements based on shift in perspective, added description of direction/trajectory	speed: s, m, f	
		Diegetic: on boat	Supported by object that is part of the shot
		cinematic	'generic' filmic movement
[Frame motion]	Frame speed. Has the action been slowed or speeded up	Speed: s, f	

Appendix E: Code Book and Trees

Appendix E.1 Code Book

INTERPERSONAL	
Agency/power	Whose voice is heard and what are the visible power relations amongst participants and with the spectator? Vertical angles. Movement. Touch.
Social distance	The solidarity and lack thereof between participants and with spectator. Intimacy
Involvement	Frontal angles signifying high involvement. Oblique, profile and 'away from' indicating the decreasing cline of involvement. Facing away becomes involved in 'co-looking'. Connects with aspects of bonding and social distance.
Appraisal: binding	The space is open/exposed (agoraphobia inducing) The space is closed/cramped (claustrophobia inducing) The space is comfortable
Appraisal: bonding	The socio-cultural nature of the space and its inhabitants supports/does not support social cohesion with the viewing spectator
Use Space / Shut-out space	(Kendon) the activated and exclusive space defined by the actions of one or more participant
Contact (combined in Mood)	Do the characters offer themselves up for observation or demand our attention? Frequency of contact spouse/child > friend > acquaintance > employer > stranger
Mood: offer	The communicative act is predicated on the giving of something (goods/services and information)
Mood: demand	The communicative act is predicated on the request for something (goods/services and information) Demand for goods and services is command
Perspective: Subjective	The spectator is 'part of the scene

Perspective: objective	The spectator observes the scene from a distance
Vis modality	Claims to types of reality. CVR claims immersive realness
Vis modality: Naturalistic Experiential Coding: high	The camera is positioned in a place/situation that could/would be occupied by a human participant.
Vis modality: Naturalistic Experiential Coding: Low	The spectator is situated in a position that would not support a human participant Low realism in temporal (e.g., retimed) Graphic interventions as barriers / hypermediacy
Modality: Modulation	Low (may) > medium (will) > high (must) modality Expression of attitude towards the truth in things (perhaps, for sure, sometimes, isn't always) Explicit, finite verb 'might', tempers the assuredness and thus status of knowledge and participant. Added modal adjuncts (was + possibly) as part of MOOD block Implicit ('is not always') and explicit ('I reckon'…'I think'…'I'm sure') Objective pretense: 'it is possible that…'
Modality: Modalization	The frequency, probability and usuality of things
REPRESENTATIONAL/IDEATIONAL	
Rank: mise-en-scene	All elements human and otherwise in the shot at any given moment
Rank: episode	Transactive clusters; sub-divisions within the mise-en-scene. Embedded narratives occur at episodic rank
Rank: figure	Discrete components, human or otherwise. Information embedded in figure denoting identity and social status
Rank: member	Sub-components of figure e.g. *body-parts* for human figure

Participant: generic	Unidentified participant that is not developed or given additional status within the text
Participant: specific	Participant is identified as part of the narrative. Participant is identified directly (presented) or indirectly (presumed)
Process: transactional	The process (in the visual field) entails the tangible causal and narrative based interaction of multiple participants. Connections are vector based
Process: reactive	
Process: conceptual	The process (in the visual field) has no vector
Pr. Conceptual: symbolic	The participant as target has attribute conferred from salient external source meaning
Pr. Conceptual: analytical	The participant has meaning based on possessive attributes has/is part of larger system of meaning
Process: material	The process (in the verbal field) is specified as a tangible causal interaction of multiple agents
Process: mental	The process (in the verbal field) involves projected thought
Process: verbal	The process (in the verbal field) involves projected speech
Process: relational	The process (in the verbal field) involves attributing a quality or identity
Process: behavioral	The process (in the verbal field) involves a self-oriented quasi-action (e.g., smiling, laughing)
Process: existential	The process (in the verbal field) involves being/becoming
Circumstance: location	Context of time
Circumstance: time	Context of place
Circumstance: manner	Means, comparison, quality, and degree
Circumstance: cause/result	Reason, purpose, and behalf
Circumstance: accompaniment	With people and things

Circumstance: Contingency	Condition (why?), default (short of), concession (despite)
COMPOSITIONAL/TEXTUAL	
Gestalt	How the elements tie-together in a holistic visual arrangement
Theme	The principal (special) *matter* within a spoken clause. The angle. The locus of viewpoint. Theme indicates modal responsibility ("I need" vs "it is required")
Info-value: given	That which is elaborated upon Often thematized elements General on left in visual
Info-value: new	That which is added to given as elaborated / effective Specific on the right Stressed words via TONICITY/TONALITY system
Info-value: ideal	Idealized generalized essence of information. Visually 'up' (lofty – idealized)
Info-value: real	Localized actual matter Visually 'down' (down to earth, grounded – real)
Info-value: Centre/margin	Non-binary, concentric informational arrangement The centre 'holds' the periphery. Importance is centre weighted.
Info-value: Front/back	Bodily determined, front is sensorially/semiotically rich. Bodies and objects are viewed frontally with their back-parts hidden (functional)
Salience	Visual 'weight' – foreground, overlapping and large (close or inherent scale) Figure weight / episodic weight (balance) The component of the visual field that is prominent, variant (figure) against an invariant field (ground).

	TONICITY in the phonological dimension (stressed words)
(sub)Framing	Abstract shapes within the mise-en-scene that outline, point to, or contain elements of importance. Lines and shapes that hold and separate participants. Disconnection and contrast, connection and integration. (sub)Framing
Rhythm	Coherence and structure in time and space The 'beat': action ~ reaction The accumulation of patterned resources that emphasize and draw interest. The repetition of words and phrases. Musical rhythmic and prosodic intonation Visual (spatial) patterns and (temporal) repetitions
Reading path	Plausible reading path moves from most salient to less salient visual elements. Connects with given-new paradigm
Cohesion: Identity tracking	Resources used in introducing participants and keeping track of them in the text. Connects with participant identification int eh interpersonal dimension. Includes the narrativized spectator. Music (leitmotifs) used to cohere identity.
Cohesion: Int. Negotiation	Resources of exchange
Cohesion: Idea. Negotiation	Recursively construing and communicating institutional activity (knowledge/reality)
Conjunction: Logical: transition: fade	Visual transition between shots fading from/to black, signals punctuation between phases/sub-phases
Cohesion: Logical: transition: dissolve	Visual transition between shots mixing between shots, signals move to new (contemporaneous/non-contemporaneous place/time)
Cohesion: Logical: transition: cut	Visual transition between shots with hard cut.
Cohesion: Logical: addition	Adds new information between clauses, shots and intersemiotically
Cohesion: Logical: elaboration	Distills by restating, summarizing information for emphasis ("in other words")

Cohesion: Logical: enhancement	Provides circumstantial features such as time, place, cause/reason, condition, result.
Cohesion: Logical: logical	Causal, conditional, comparative connection between elements
Cohesion: Logical: Projection	Projection of thoughts or speech
Paraphoning	Patterns of segmented speech based on situational requisites.

Register Coding	
Code	**Description**
Field	
The semiotic and social activity (Halliday & Hasan in Shore, 2014) " SOCIO- SEMIOTIC PROCESS ('that which is going on') and the PHENOMENAL DOMAIN ('subject matter')" (Matthiessen in Ventola and Guijarro, 2009, p.28) First order (social (Interpretable via the ideational coding)) and second order (semiotic) Semantic Domain: The overarching domain or discipline (high level ideational content) "taxonomies – groupings of people, things and processes; these taxonomies in turn distinguish one fi eld from another" (Martin and Rose, 2008, p.14)	
Phenomenal Domain	
Phen.D: Process types	Concrete local processes (low/mid-level) "The topic of the situation" (Eggins, 2004, p.103) Realized in ideation MF process types Identified through lexical words and nouns (Měchura, 2005)
Phen.D: Circumstance types	Locating and characterizing the processes Realized in ideation MF circumstance types
Specialization	Knowledge entry-level – how specialized is the ideal audience?
Socio-semiotic process	'that which is going on'

		Second-order semiotic field (sub-codes) Socio-semiotic types – can 'shade' into each other (Matthiessen) Italics from Matthiessen's 'Register in the round'
	Socio-semiotic process: Expounding	categorizing/explaining *contexts where natural phenomena such as cold fronts are explained to help readers or listeners as part of the construction "knowledge" about general classes of phenomena.*
	Socio-semiotic process: Reporting	chronicling/surveying/inventorying *contexts where the flow of particular human events are chronicled to help readers or listeners construct keep up with or review events.*
	Socio-semiotic process: Recreating	narration/dramatization *contexts where the flow of particular human imaginary events are narrated to achieve some kind of aesthetic effect.*
	Socio-semiotic process: Sharing	experiences/values *contexts where personal values and experiences are exchanged to help interactants relate to one another for example by calibrating their sense of moral values in a work place*
	Socio-semiotic process: ***Doing***	***collaborating/directing*** ***contexts where people are engaged in a joint social activity, using language to facilitate the performance of this activity.***
	Socio-semiotic process: Recommending	advising/inducing *contexts where a course of action is advised for the benefit of the addressee*
	Socio-semiotic process: Enabling	instructing/regulating *contexts where a course of action is modelled semiotically and made possible through instruction.*
	Socio-semiotic process Exploring	Arguing: Expositions/discussions *contexts where public values and ideas are put forward and debated*

Tenor

Participant relationships (Halliday & Hasan in Shore, 2014)

The kind of person the author and expected audience are - or are pretending to be (Měchura, 2005) "The social

role relationships played by interactants" (Eggins, 2004, p.99)

Three overarching axes: power, contact and affective involvement, which when overlaid give higher level descriptions of the situation regarding its formality (Poynton in Eggins, 2004, p.100) Intimate >>> public (Math in V-G)	
Status	Relative status "terms of address; who gets to choose the topic; who gets to choose who speaks (Měchura, 2005) Continuum of equal to unequal power (reciprocal/un-reciprocal) (Eggins, 2004, p.100) Who dominates and who defers? (Martin and Rose, 2008) Mood component (all contextually determined) - **Statement makers:** holders of worthwhile information. Assessments infer authority and competence. - **Questioners:** 'needing' answers from authority holders - **Order givers:** controllers of others' behaviour Who is in a (visual) position of dominance?
Standing	Expertise and authority Credible secondhand sources of information Demonstrating expertise In a position to praise/criticize

Agentive Role	The inferred social role of the participants and spectator (Hasan below)
	
	- Acquired/inherent.
	- Civic:
	    ○ Office: supervisory/negotiated
	    ○ Status: rights, expertise, achievement
	- Familial
	- Reciprocating: equal, complementary
	- non-reciprocating

Stance	Does the author/participant allow for interpretation or disagreement
Stance: Attitude	Negative/positive meanings expressed Explicit, assumed and triggered attitudes. - Author uses negative lexis to imbue the text with an attitudinal force - Neutral / Evaluative ('feature' versus 'weakness')

	- Evaluative adjectives
	- Agency/affectedness: structuring as actors (agents) and goals (affected)
	    o The police shot the demonstrators (transactional actors)
	    o Shots were fired by the police (non-transactional actors)
	  - Nominalizations: removing actors ("Tuesday saw much shooting")
	- Ergative: "the shooting continued"
	-

Stance: Modality	- Epistemic: is the content expressed as *true*
	      o Will, would, must, may…be true
	- Deontic: how much obligation is placed on the spectator to act, given the message of the text
	    o Must, have to, ought to, be supposed to…do that

Social/personal Distance	The 'construction of biography' (Hasan) within fields of activity
	
	AKA 'contact'
	
	Visual Field:
	
	  - Proxemics
	
	Implied frequency and range of contact between text participants and spectator
	
	Has Goffman connection (Facework)
	
	The projection of *persona* (Měchura, 2005)
	
	'Familiarity'; formal, informal vocabulary (Měchura, 2005)
	
	  - Colloquial / formal language
	  - Contractions
	  - Close: Ellipsis suggests shared knowledge. Being 'in on' things.
	    o Exophorics: (Hasan in Lam 2016)
	  - Distant: 'tentativeness/uncertainty' (In modality)
	The creation of solidarity in non-interactive texts
	
	Real/deliberate im/personalization. How much attention is drawn to the writer (Měchura, 2005)
	
	  - Personalization – personal pronouns
	  - Rhetorical questions and imagined questions coming from the reader
	  - Shared context 'here' 'now' "to draw the reader in" (Měchura, 2005)
	Poynton's 'proliferation and contraction'. Proliferation: how open can people be – how much can interlocutors speak about. Contraction: how much work is required to exchange meanings (in Martin and Rose pg. 13

Affective involvement	Implied closeness of relationship between text participants and spectator
Mode	
Axes of Mode	

- Medium and Channel
  - **Medium:** spoken > written
  - **Channel (sensory):** aural, visual, tactile, olfactory, gustatory
  - 'Operating in the same spatiotemporal realm […] process instances in real-time' (Matthiessen)

CHANNEL

'The semiotic construction of communication technology…representing the semiotization of the affordances of the material channel'

'The bandwidth of semiosis between 'speaker' and 'addressee'. Minimal (not present) > maximal (colocated)

DIVISION OF LABOUR (Matthiessen) – which modes do the 'heavy lifting'

TEMP PHONOL.	PARATONES: Esser (1988: 26, 29) cautions that those who make their living through oral renditions of scripted material are more sensitive to the communicative value produced by high onsets and low terminations.
Axis of medium: Language situations	Archetypal language situations Spoken and written, being the extremes of usage types All else is in between
Language sit: Spoken	- Face-to-face - Interactive - Language as action - Spontaneous - Casual **Visuals:** **Text:** **Dynamically linked clauses:** 2 + clauses clauses connected explicitly (logical: because) **Human actors:** personalized (human) participants (I, my kids)

	action processes with verbs (handed in, got sick) turn-taking organization context dependent dynamic structure interactive staging open-ended spontaneity phenomena (false starts, hesitations, interruptions, overlap, incomplete sentences) Everyday lexis non-standard grammar grammatical complexity lexically sparse

Language sit: Written	**Condensed sentence**: one longer clause
	
	depersonalized (thematized) 'I' replaced with abstract 'reasons'
	
	Action process becomes relational process (was) Logical relation (because)
	
	becomes noun (the reason)
	
	Actors (kids) become possessors ((illness) *of* my children)
	
	Nominalization: Verbs (hand in, get sick) are nominalized (submission, illness)
	
	monologic organization
	
	context independent synoptic structure
	
	   - rhetorical staging
	   - closed, finite
	final draft (polished) / indications of earlier drafts removed prestige lexis
	
	standard grammar grammatical simplicity
	
	lexically dense

Monologic/dialogic	- can interlocutors see each other? - Is intended to appear as if dialogue CAN happen

	Whether interlocutors can hear and see one another (aural and visual feedback) and the imminence of a response (immediate or delayed). M+R 15

Appendix E.2 Code Tree: interpersonal coding

interpersonal/interactive

- mood/contact
  - lxg-imperative
  - lxg-interrogative
  - lxg-declarative
  - vis-offer
  - vis-demand lxg-
- modality
  - modulation lxg-
  - modalisation tonality
  - vis-naturalistic-coding
    - vis-high
    - vis-low
- vis-perspective
  - vis-vertical-power
    - above houlder
    - slow
    - t
  - vis-liminal
  - vis-moving
  - point-of-view
    - objective subjective
- vis-distance
  - extreme public
  - personal
  - intimate
    - low medium
    - high
- vis-
- involvement
- vis-use-space

vis-appraisal
- bound
- unbound

insecure secure

b-insecure ⊣ b-secure

Appendix E.3 Code Tree: Ideational/Representational Codes

indeational-representational
- visual-rank
  - mise-en-scene
  - episode
  - figure member
- process
  - causative-reactive
  - existential
  - behavioural
  - relational
  - verbal mental
  - -material
  - vis-transitivity
    - action
      - transactional
      - non-transactional
    - reaction
  - conceptual
    - relational-attrib
    - symbolic-attrib
- participant
  - status
    - specified
      - presented
      - presumed
    - generic
  - role-(all-sfl-role-types)
- circumstance
  - projecting
  - elaborating
  - extent
  - matter role
  - reason
  - contingency
  - accompaniment
    - with-generic with-specified alone
  - cause/result
  - manner time
  - place
- conjunction
  - transition
    - cut
    - add-dissolve fade enhancement elaboration
    - extension
  - expansion

Appendix E.4 Code Tree: Textual/Compositional Coding

textual-compositional
- salience
  - gradual
  - immediate
- framing
  - plane
  - container
  - stage
- geometric
  - negative-space
  - geometric-perspective
  - triangular-composition
- reading-path
  - leading-lines
  - intershot-rp
    - start-probability-frame
    - end-probability-frame
- tone
  - marked-tonality
  - unmarked-tonality
- information-value
  - given
  - new
  - vis-ideal/above
  - vis-real/below
  - vis-centre/margin
  - vis-front/back
  - theme
    - textual
    - interpersonal
    - topical

Appendix F: Field Breakdown

Appendix F.1: Socio-semiotic Field Types

Existing typological category	Clarified usage of typological category
Reporting	
*"reporting on our experience of particular phenomena — **chronicling** the flow of* *particular events (as in historical recounts or news reports), **surveying** particular* *places (as in guide books) or **inventorying** particular entities (as in catalogues)"*	**Chronicling**: reporting <u>temporal</u> events as part of the flow of observable history **Surveying**: to report/examine/describe (<u>nontemporal</u>) entities for their observable external features. This is extended to people, places, and things.
Expounding	
"expounding our experience of classes of phenomena according to a general theory *(ranging from commonsense folk theories to uncommonsense scientific theories* *— either by **categorizing** (or* *"documenting") these phenomena or by **explaining** them"*	**Categorizing** entities for their attributes. Unlike surveying, this shades into subjective and conceptual descriptions involving comparisons **Explaining** entities through causation and argument
Recreating (no change required)	
*"**recreating** our experience of the world imaginatively, that is, creating imaginary worlds having some direct or tenuous relation to the world of our daily lives — recreating the world imaginatively through narration and/ or through **dramatization**"*	
Exploring	
*"**exploring** our communal values and positions, prototypically in public — either by* ***reviewing** a commodity (goods-&-services) or by **arguing** about positions and ideas"*	**Exploring** Will include subjective 'pondering' in the real or imagined presence of a listener
Advising	

*"**recommending** people to undertake some activity, thus very likely foreshadowing a* *'doing' context — either by **advising** them (recommendation for the benefit of the* *addressee, as in consultations) or **inducing** them (promotion: recommendation for* *the benefit of the speaker, as in advertisements);"*	**Advising** (as sub-type of recommending): will include advising the listener of the attributes of a phenomenon 'for the listener's sake', as well as recommending an action
Sharing (no change required)	
"sharing our personal lives, prototypically in private, thereby establishing, maintaining and negotiation *personal relationships — sharing our personal experiences and/ or sharing our personal values"*	
Doing (no change required)	
"contexts where people are engaged in a joint social activity, using language to facilitate the performance *of this activity"*	

Appendix F.2: Subject Matter: Congo VR

The Production and attached spectator		DRC Geography		Politics		Corruption	
you	7	DRC	5	their own leaders	1	his empire of riches	1
The team and I	5	Kinshasa	2	Mobuto's private palace	2	Muhammad Ali	1
Alastair Leithead	2	gBadaliter	2	the people	3	pink champagne	1
BBC Africa Correspondent	1	Buyanger Village	1	Mobutu	3	his bed	1
(V) BBC News logo	1	One of the … cities	1	totalitarian	1	below the floor	1
(V) drone	4	The capital	1	Mobuto's regime	1	the ruined palace of president Mobuto Sense Seko	1
(V) Alastair Leithead.	9	sprawling mass of more than ten million people	2	Traditional chief of zambia	2	a long marble table and a gigantic chandelier	1
(V) CLUSTER #2 BBC CREW	2	two thirds of this country	2	KPWATA FYFY (+)	2	His grand dining room	1
(V) CAMERA OP]	1	under 25	1	a flicker of hope	1	blessing	1
(V) BBC WOMAN	1	(V) JUNIOR	2	independence	1	Beautiful but troubled place	1
		(V) PYGMY ELDER	1	Mobuto's town	1	The rich and famous	1
		(V) MOBUTU DAUGHTER	1		**18**	paranoid	**2**
		(V) Generic	18	**DRC Culture**		riches	4
	33	(V) PRINCESS MAMICHO	1			Picasso's paintings	1
Historic conflict		(V) WOMAN IN RED	1	a sapeur	5	very beautiful	1
At the hands of the colonizers	1	my mother (Pr Mamicho)	3	pygmies	5	(V) large swimming pool	1
history	1	junior	2	a good luck ritual	1	(V) font-like structure	1
Europeans	1	the energy and brash confidence of the capital	1	fashion	1	(V) large dilapidated building	1
militants	1	one of his daughters	1	an artist	1	(V) cross-shaped pool of water	1
troops	1	his family	1	people do exciting and surprising things	1	(V) large swimming pool	1
a coup	1			(V) CLUSTER #1 DANCERS	1	(V) font-like structure	1
their brutal Belgian rulers	2			(V) TRIBESMEN	2	(V) large dilapidated building	1
the wounds of the past	1				**17**	(V) cross-shaped pool of water	1
the difficulties of living here	1	**(minus generic p = 31)**	**49**	**Resources**			**27**
Ebola	1			Its mineral wealth	2		

war	2			our /heritage	1	
poverty and conflict	2	**Travel/Exploration**	1	Rubber	4	
curse	1	the railway	1	animals and useful plants..	1	
still raw	1	This' (tain travel)		that white sap	1	
to burst free	1	very happy (that you have visited)	1	clothing and shoes and elegance	1	
to make a better life	1	(travel is never) easy	1	(V) Construction vehicles	1	
future	1	travel	1	(V) Aerial vista of barren, sandy terrain	1	
(V) SOLDIER	1	from exploring…Congo	1	(V) MAN DIGGING	1	
(V) [GFX] black and white still of Mobutu's face	1	another way around	1	(V) MAN BOW/ARROW	1	
(V) CONGO A TROUBLED PAST	1	a great adventure	2	(V) MAN GREEN SHORTS	1	
our own sisters	1	Mighty Congo river	2	(V) CLUSTER #1 MEN LIGHTING FIRE	1	
		Congo river rapids	2	(V) fire on ground	1	
		the train to Kinshasa	1		**17**	
		(V) landmass, water (sea meeting river delta)	1			
		(V) Aerial vista of forest treetops	1			
		(V) 5 persons towards rear of boat seated	2			
		(V) Blue train with bright red/yellow stripes at platform	1			
		(V) taxi-car immediately on front	1		32	
		(V) MOTORCYCLIST	1			
		(V) Fast moving white-water river	1			
	23					
			20			

Appendix F.3: Subject Matter: Mamie's Dream

Social recreation & play		Family		Personal conflict		Celebration	
(V) CLUSTER #2 (playing under tree)	1	M family	1	daughters father	4	(V) CONGREGATION	1
(V) CLUSTER #1 (playing under tree)	1	father	7	people / community	4	(V) dense crowd of pupils	1
(V) DANCING PUPILS	1	mother	1	belongings	1	(V) MAN AT LECTERN	1
(V) Boys playing	1	SL families	1	painful	1	(V) BAND	1
(V) football	1	stepother	2	difficulties	1	(V) DANCING MAN	1
(V) READERS	1	parents	3	FGM	1	**Celebration**	**5**
(V) CROWD (watching football)	1	(V) daughter 7	2	pregnancy	1		
(V) woman cooking	1	his family	1			**Authority figures**	
(V) a woman in the centre of the area cooking	1	**Family**	**18**	**Personal conflict**	**13**	(V) A group of adults (on stage 18)	1
Social recreation & play	**9**					(V) MAN (w/dancing girls)	1
						headmaster	2
Gender		**Education**	**2**	**self-determination/actualization**		male teachers	1
(V) Specified female pupil	4	LAP	3	comm leader / role model	2	(V) adult sits at a table	1
		school	1			(V) man (boat)	1
womanhood	1	(V) PLAN INTENATIONAL logo	1			Bundu	2
		(V) A circular structure	2	dream	3		
girls (gen)	5	secondary school	1	nurse	1	**Authority figures**	**9**
one girl	5	students	1				
boys	1	6 subjects	3			**self-determination/actualization**	**7**
Gender	**16**	(V) classroom pupils	1	proud	1		
		(V) Classroom of adult women	3			**Mamie**	
		(V) Blackboard	3			Mamie ('I')	33

5

			(V) Generic male teacher	3			(V) Mamie	17
			teacher (Mamie)	3				
			Education	**22**			Mamie	**33**

Appendix G: Lexical Density Analysis

Mamie	Spoken under 45	Fiction writing 45-55	Expos 50-60	Technical 60-70
1 IN BOAT	46.67		46.67		
2 ON BRIDGE	72.73				72.73
3 KIDS PLAYING FOOTBALL	61.54				61.54
4 UNDER TREE IN YARD	50		50		
5 STANDING IN DOORWAY	35.71	35.71			
6 #1 SCHOOLROOM	42.42	42.42			
7 SITTING ON BED	38.64	38.64			
8 IN MARKET	37	37			
9 SITTING UNDER SHELTER	36.36	36.36			
10 ADULT LEARNERS	50		50		
11 Mamie ASSISTING	42.11	42.11			
12 Mamie TEACHING	50		50		
13 OPEN AREA	51.85		51.85		
14 CHURCH					
15 HOLDING HANDS	44.12	44.12			
16 GIRLS PLAYING	56.76			56.76	
17 SERIOUS CHAT	44.83	44.83			
17 SERIOUS CHAT (Diegetic)	63.16				63.16
18 DANCING CLAPPING					
19 RIVER	50		50		

| | 47.15625 | | 47.15625 | | |

CONGO		Spoken	Fiction writing	Expos	Technical
1 DRONE OVER DELTA	57.89			57.89	
2 DRONE FORESTRY	46.67		46.67		
3_DRONE SANDY LANDSCAPE	55		55		
4_PRESENTER IN BOAT (Diegetic)	44.4	44.4			
5 DRONE RIVER TITLE					
6 DRONE RIVER RAPIDS	52		52		
7 TRAIN STATION #1					
8 TRAIN INTERIOR WITH AL	65.22				65.22
9 TRAIN INTERIOR WINDOW	30	30			
10 TRAIN STATION #2	63.64				63.64
11 KINSHASA TRAFFIC	63.16				63.16
12 DRONE ROOFTOPS	52		52		
13 JUNIOR	55.56			55.56	
14 YARD PRINCESS MAMICHO (Diegetic)	38.71	38.71			
15 DANCEFLOOR	56.25			56.25	
16 DRONE STATIC FORESTRY	57.14			57.14	
17 VILLAGE TRIBE	47.37		47.37		
18 FOREST HUNT	50		50		

19 FOREST HUNT SHOT #2	57.14			57.14	
20 FOREST HUNT SHOT #3	43.48	43.48			
21 PYGMY MAN SPEAKS (Diegetic)	50.91		50.91		
22 VILLAGE - UNDER CANOPY	50		50		
23 DRONE TOWARDS MOBUTU'S RUINS	50		50		
24 MOBUTU'S RUINS	67.65				67.65
25 MOBUTU'S DAUGHTER (Diegetic)	52.63		52.63		
26 PALACE RUINS #1	50		50		
27 PALACE RUINS #2	62				62
28 PALACE RUINS #3	46.81		46.81		
29 DRONE SUBURBS	66.67				66.67
30 MILITARY VEHICLE	52.67		52.67		

53.03464286

Mamie's Dream
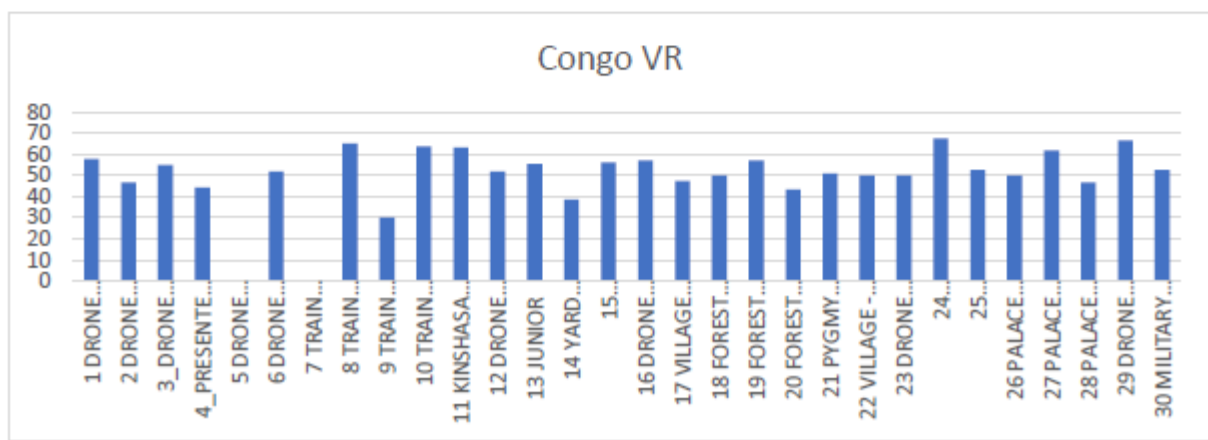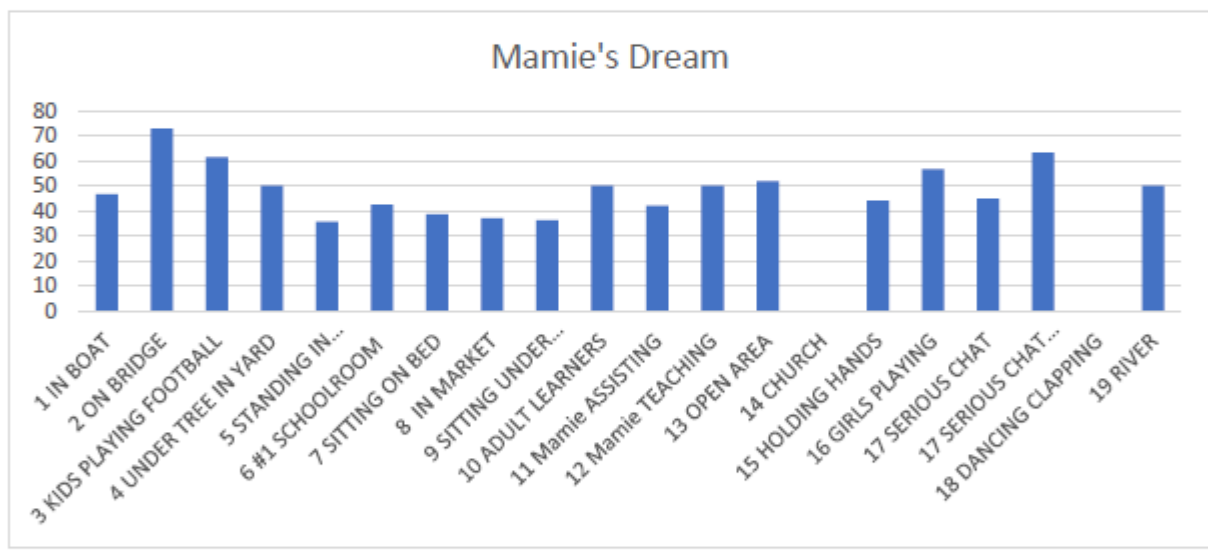
Congo VR

Appendix H: Detailed Cohesion Analyses

Appendix H.1 Congo VR Cohesion

Shot #16

Location: The shot begins with a re-orientation of comprising an anaphoric reference to 'the Capital' from the previous phase, followed by a cohesive tie between 'far' (from the Capital) co-hyponym as units of distance, with 'deep' (in the forest). A cross-modal parallel is formed between 'deep in the forest' and the visual forest scene. 'The forest' is presumed deictic reference (assumed to exist) or presumed and anaphorically retrieved from an earlier phase ('with the largest stretch of rainforest' C#2).

> [presumed] + {within shot [anaphoric] + [bridging: hyp.]} + [cross-modal parallel] + [cataphoric]

Suffering: 'Wounds of the past' is a presumed generic entity that commences this chain, realized cataphorically through ties shot 21 with 'suffering'.

> [presumed] + [generic] + [cataphoric]

colonizers: The circumstance '[...] Belgian colonial times' is presumed and cataphorically tied with a chain beginning in shot 19 ('Brutal Belgian rulers', 'the colonizers', 'white people', the Belgians).

> [presumed] + [specific] + [cataphoric]

The Pygmies: 'Pygmies' is presumed, generic in the lexis and the sounds of the chanting tribesmen is heard establishing a prelude/cataphoric tie with the next shot (also *meronym* when voice is considered 'part' of those visualized in the next shot).

> [Lex: presumed] + [generic] + [diegetic sound: prelude/cataphoric] + [cross-modal reference]

Shot #17

Location: The forest here is realized anaphorically through bridging tie (meronym) with the previously presented 'forest' (including the clay road), here seen at ground level as part of the village. The village is referenced cross-modally by deixis, 'this is – Buyanga Village'.

[presented] + [bridging: meronym] + [cross-modal deictic reference]

Pygmies: the previously presumed pygmies (as diegetic prelude) are presented here as visually salient and specified cross-modally by repetition. They are referenced cross-modally through deixis 'they(re)' and 'them'.

[presented] + [lex: mono-modal repetition] + [vis: salience] + [cross-modal deictic reference]

Ritual/Hunt: a chain of ritual-hunt components is presented cross-modally, through deictic reference ('they're performing...hunt'), salient tribesmen and diegetic chanting.

[presented] + [vis: immediate salience] + [cross-modal deictic reference]

'We' (BBC team): 'we' are presumed anaphorically from earlier shots ('The team and I'). The BBC team or part thereof are visually repeated from shots 4, 8, 15. 'We' and the visually salient BBC are thus realized cross-modally.

[presumed] + [cross-modal deictic reference] + [anaphoric]

Shot #18

Location: the forest has a bridging tie of co-meronym with the visual ground-level view from previous; we see the forest at the same level of detail and at the same level.

[presumed] + [bridging: co-meronym] + [anaphoric]

Pygmies: pygmies are presumed, referenced anaphorically as 'they(re)' while also referenced deictically between the voice-over and the visual, as well as their dialogue audible in the diegetic audio track. They are presumed visually through visual repetition and further specified cross-modally in the diegesis.

[presumed] + [vis: repetition] + [specified] + [cross-modal salience (diegetic)] + [cross-modal deictic reference]

Ritual/Hunt: 'looking' is a presumed anaphoric tie of collocation with 'forest hunt' ('hunting' requiring 'looking'). The inter-visual cohesion here is of repetition, where each tribesman follows the same movements and trajectory, emerging from and returning to the open tree-trunk. Parallel processes enable cross-modal reference.

2

[presumed] + [lex: bridging: collocation] + {within shot [repetition] + [anaphoric]} + [anaphoric] + [cross-modal parallel]

Shot #19

Forest: the bridging tie of co-meronym continues here where the forest is seen at the same level of detail but in a different location.

[presumed] + [bridge: co-meronym] + [anaphoric]

Pygmies: the pygmies are repeated visually. A verbal reference 'they(ve)' creates an anaphoric tie to the cross-modally specified 'pygmies'.

[presumed] + [visual repetition] + [lex: reference]

Man digging: a salient component (man digging) is presented visually with immediate salience. He has a meronymic tie as part of the group.

[presented] + [generic] + [vis: immediate salience] + [bridge: meronym] + [anaphoric]

Pygmy Elder: The pygmy elder (presented cross-modally in successive shots) is referenced cataphorically here through diegetic audio prelude. He is visible also

[presumed] + [repetition] + [diegetic sound: prelude/cataphoric] + [vis: prelude/cataphoric]

Ritual/hunt: The hunt is presumed, referenced anaphorically from the previous shot partly through the continued diegetic sound of hacking. 'hunted' is a repetition from previous mentions of 'hunt' and collocative from 'looking'. 'hunted' is further tied with 'like this' through connection of *comparison*, which sets up another deictic tie with the visualized pygmy who is thus perceived as engaged in a hunt-related process.

[presumed] + [lex: collocation / comparison] + [diegetic repetition] + [anaphoric (esphoric)] + [cross-modal: deictic reference]

Rubber: 'something in the forest worth a fortune' is presumed cataphoric reference tying with verbal and visual components in next shot.

[presumed] + [cataphoric] + [mono (lex) and cross-modal reference]

Colonizers: 'their Brutal Belgian rulers' is yet to be presented and specified. It remains presumed through 'their' and cataphoric to the more explicit mentions made later. The tie is with 'the colonizers' and as such one of collocation.

[presumed] + [cataphoric] + [bridging: collocation]

Shot #20

- Time: 'a century ago'
- Forced to search – hunt – looking?

Forest: the bridging tie of co-meronym continues here where the forest is seen at the same level of detail but in a different location.

[presumed] + [bridge: co-meronym] + [anaphoric]

Pygmies: the pygmies as a bloc are repeated visually, as similar standing and observing position. 'the pygmies' is presented verbally but remains lexically generic (he does not say 'such as this tribe') in contrast to the visual specification through visual repetition. Crossmodally they are bridge from verbal to visual as co-hyponym/meronym[46]

[presumed] + [visual repetition] + [lex: generic] + [vis: specific] + [cross-modal bridge hyponym/meronym]

Man Digging: the man, salient in the previous shot has no salience here and re-merges to the tribe component. The tie is roughly that of part-whole bridging (meronym).

[presumed] + [vis: bridge: meronym] + [anaphoric]

Pygmy Elder: similarly to the man digging, this component emerges as part-whole bridging meronym tie and although cross-modally salient (visuals and diegetic dialogue), he is not fully presented and cataphorically referencing the next shot. A tie of visual synonym is also plausible here when he is roughly matching the position and activities of the digging man in the previous shot.

- (To tribe) [presumed] + [vis: bridge: meronym] + [diegetic: repetition (from prelude)] + [anaphoric]

---

[46] 'Hyponym' when the pygmies are considered a class/race of people, 'meronym' when considered as acontinuous physical extension of the DRC pygmies around them and before them, in time.

4

- (To previous man) [presumed] + [bridge: synonym] + [anaphoric]
- (To himself in shot 21) [presumed] + [repetition] + [gradual salience] + [cataphoric]

Ritual/hunt: an issue emerges here where the knife action is cohesive with previous hunting activities (as collocation – knives are used to hunt) but is in fact the process of extracting rubber. The process component here becomes fused with the *rubber* reference chain, that begins here with significant ties thereafter. The hunt chain terminates here.

[presumed] + [collocation] + [anaphoric]

Rubber: 'rubber' is presented verbally as a generic entity and referenced as 'it' and 'that white sap oozing out of the wild vines', which also contains a deictic cross-modal reference to the pygmy elder cutting into a vine, extracting rubber.

[presented] + [specified cross-modally as deictic reference]

BBC: AL is presumed, visibly salient in the group, tied anaphorically (meronym) with previous 'we'.

[presumed] + [vis: immediate salience] + [anaphoric]

Shot #21

Forest: the perspective is slightly different, but the location is the same as in the previous shot and is tied by visual repetition.

[presumed] + [vis: repetition] + [anaphoric]

Suffering: suffering is presented with verbal ties of repetition in the verbal track, 'the suffering' and 'suffered' both hyponyms with the earlier 'wounds of the past', under the general class of suffering.

[presumed] + [lex: repetition] + [bridge: hyponym] + [exophoric]

- Time: 'started' coll. 'into today'

Colonizers: In the pygmy elder's speech (via translation) 'the colonizers' are presumed a generic entity that is tied through hyponym to 'the white people', under the general class of 'European', which is then bridged similarly with 'the Belgians'.

[presumed] + {*lex within shot* [bridging: hyponym] + [cataphoric]}

Pygmy elder: the pygmy elder is presented here visually with increased salience and cross-modally through his translated diegetic speech as well as the graphic of his personal name *Ilungama Ayanda*

      [cross-modally presented] + [gradual salience (increased from previous)]

Pygmies: 'our own sisters' and 'our ancestors' (referenced as 'them') are co-hyponyms of 'pygmies', referencing anaphorically. Visually the pygmies are repeated anaphorically in roughly the same peripheral stance.

      [presumed] + [bridge: co-hyponym] + [anaphoric]

- Forced to collect > collecting is in rubber chain and hunting chain
- Rubber: 'them' ref from previous CM Parallel / VIS vine leaf meronym with previous Vine

Rubber: 'rubber; already specified is repeated and referenced pronominally ('it') in his verbal chain. Rubber as a chain of component ends in this shot. [presumed] + [generic]

Heritage: as part of his speech, 'heritage' is introduced with cataphoric collocation ties to 'left the people', but in the context of the next phase. It is a prelude of sorts and a skillful way of implicitly threading the phases together.

Reference chains: River; Time; Dream; Mamie; Parents; Boys; Girls; School; Men; Play area; Society; FGM

Shot #0 (Titles)

Dream: graphic title 'Mamie's Dream' presents the dream, specified as a particular instance of a dream (Mamie's)

> [presented] + [specified]

Mamie: title presumes 'Mamie'.

> [presumed]

Shot #1

River: The river is salient cross-modally as visual backdrop and as gradually salient diegetic audio (bird sounds and water heard with increasing volume).

> [presented] + [diegetic: gradual salience]

Dream: the dream is specified verbally as 'my dream'. Reiterated cross-modally. [presumed] +

> [anaphoric] + [cross-modal reiteration (from previous shot)]

Mamie: woman is presumed visually to be Mamie, specified through strong immediate salience. Presumed in the lexis (I, I)[47].

> [presumed] + [specified] + [immediate salience]

Father: 'my father' is presented and specified verbally (has no visual).

> [presented] + [specified]

---

[47] I am not assuming a projection here, in which case Mamie would be presented as owner of dream

Parents: 'My father and mother' are presented as unit, and as superordinate meronym for 'father'.

[presented] + [specified] + [anaphoric] + [bridge: meronym]

Girls: '(Educating) a girl' is presented in her father's clause as generic. '(when I was) a little girl' is presented and specific, while anaphorically tying back as hyponym ('a girl' > 'this girl').

[presented] + [generic] > [presented] + [specified] + [anaphoric] + [bridge: sub-hyponym]

School: 'educating (a girl)' (picked up later across modes, has no further tie in this phase)

Shot #2

River: the sounds of the river carry across the edit, as presumed esphoric/anaphoric reiteration. The river is reiterated visually or as a meronym tie (previously was part of what we now see as 'whole' river.

[presumed] + [anaphoric] + [bridge: meronym] + [cross-modal] + [bridge: synonym]

Mamie: She is further specified through visually repetition and strong salience. She also references herself as 'I'

[presumed] + [anaphoric] + [repetition] + [monomodal (lex and image separately)]

Father: 'my father' repeats earlier iteration and is also a tie of subordinate meronym from 'my father and mother'

[presumed] + [anaphoric] + [[bridge: meronym]

Girls: 'girls (they just…)' is generic and has hyponym tie with previous 'a little girl'. A broader tie of hyponym is available cross-modally with the visually salient Mamie, under the class of *female*

[anaphoric] + [bridge: hyponym] + [monomodal] + [cross-modal (with Mamie)]

Shot #3

Parents: 'parents' are presented and tied lexically as collocation/synonym with 'father and mother'. They are repeated in the shot also as 'my parents'. Additionally they are explicitly tied

synonymously with 'many families in Sierra Leone' and form a tie of hyponym (class: SL families) with that genericized unit.

[presented] + [specified] + [anaphoric] + [bridge: collocation/synonym] > [presented] + [anaphoric] + [repetition]

Boys: 'their boys' is tied cross-modally with the immediately salient footballing boys, as bridge of co-hyponym.

[presented] + [immediate salience] + [cross-modal specification] + [bridge: co- hyponym]

Girls: 'their girls' is presented as specific instances of girls, tying as sub-hyponym with previous 'girls'.

[presented] + [specific] + [anaphoric] + [bridge: sub-hyponym] Recreation-area: the

footballing and spectating areas are presented.

[presented]

FGM: '(enter into) *womanhood'* is presumed, ties cataphorically with the following clauses as bridging tie of synonym.

[presumed] + [generic] + [cataphoric] + [synonymy]

Shot #4

Mamie: Mamie gains gradual salience, swapping it with the girls who through movement and size have immediate salience at the onset of the shot. One instance of 'I' continues her verbal chain.

[vis: presented] + [gradual salience] + [anaphoric] + [repetition]

Girls: The playing girls are immediately salient and are in a chain already from verbal 'girls' (as well as a loose tie of hyponym with Mamie). They are at the same time in an antonymic tie with the playing boys in the previous shot. Play is the commonality that contextualizes the cohesive relationship.

[presumed] + [anaphoric] + [immediate salience] + [cross-modal: bridge: co-hyponym (girls)] +

9

[cross-modal: bridge: antonym (boys)]

Recreation-area: the play area here is co-hyponym with the previous shots locative circumstances.

[presented] + [bridge: co-hyponym]

FGM: 'it' forms a cataphoric referencing tie with FGM. FGM is presented and specified verbally.

[presumed] + [cataphoric] + [reference] > [presented] + [specific]

Shot #5

Mamie: three instances of 'I' continue her verbal chain. Visually, she has a degree of salience but less than the previous shot.

[vis: presented] + [anaphoric] + [repetition]

FGM: FGM can only be recovered here implicitly as it is ellipsed in her verbal chain i.e., 'I refused (it)', 'I was scared I might die (from it)'.

[presumed] + [anaphoric] + [ellipsed reference]

Appendix H.3: Identity Chains from Congo VR, Shots 16-21

	Loc. Forest	Pygmies	Pygmy Elder	Man Digging	Ritual/Hunt	Rubber	Colonizers	BBC Team	Suffering
16	'the forest' ↔ [v]	[Sound: men chanting]						'we' ↔ [v]	'wounds of the past'
17	[v]	[v + dieg]			'performing…hunt'				
		'they('re)'…'them'							
18	[v]	[v + dieg]			'looking' ↔ [v]				
		they('re)'							
19	[v]	[v + dieg]		[v]	'hunted' ↔ [v]	'something in the forest'	'Brutal Belgian rulers'		
		'they'							
				[v]					
20	[v]	[v + dieg]	[v]		[v: knife]	'Rubber'	'the colonizers'		
		'the pygmies'							
						'that white sap' ↔ [v]			
21	[v]	[v + dieg]	[v + dieg]			[v: leaf]			'the suffering…suffered'
		'our own sisters' ↔ [v]							
		'our ancestors'							

Solid arrow = explicit tie ——→
Dashed arrow = bridging tie - - -→

Appendix I. Illustrations of Conjunctive Analyses

Shot	CRU	Visual	CRU	Verbal (incl. explicit conjunctions)
#1	V1	Mamie in boat	1	My father said
	V2	Man rowing	2	educating a girl is a waste of money
			3	When I was a little girl, I told my father and mother,
			4	my dream was to become a teacher
#2	V3	Mamie in bridge	5	My father said, "
	V4	River	6	girls, they just get married and move away"
#3	V5	Footballers	7	My parents loved their boys
	V6	Crowd	8	(+) Like many families here in Sierra Leone, my parents wanted their girls to join the Bundu society and enter into womanhood
#4	V7	Mamie	9	(+) I heard
	V8	Girls playing	10	it's painful
	V9	Girls slowed down	11	(=) They blindfold you and cut your clitoris
			12	(=) It's called FGM

Shot	CRU	Visual	CRU	Verbal
#16	V1	Forest	1	Far from the capital, the wounds of the past are still raw
			2	(=) Deep in the forest, pygmies live a life where little has changed since Belgian colonial times
#17	V2	Tribesmen	3	(+) This is Buyanga village.
	V3	BBC crew	4	(+) they're performing a good luck ritual before a forest hunt.
	V4	Village	5	(+) and we're going with them
#18	V5	Tribesmen	6	(+) They're looking for animals and useful plants along the way
	V6	Forest		
#19	V7	Man digging	7	(x) They've always hunted like this
	V8	Tribesmen	8	(x/+) but a century ago, their brutal Belgian rulers discovered something in the forest worth a fortune
	V9	Forest		
#20	V10	Pygmy elder	9	(=) Rubber
	V11	AL	10	(=) It's that white sap oozing out of the wild vines.
	V12	Tribesmen	11	(+) The pygmies were forced to search for them, and to collect it.
	V13	Forest		
#21	V14	Pygmy elder	12	(=) Our own sisters suffered at the hands of the colonizers collecting rubber
	V15	Tribesmen	13	(+) rubber was our heritage,
	V16	Forest	14	(+) and we did not benefit from it
			15	(= 12) The suffering started with our ancestors
			16	(+) and it continues into today
			17	(= 12) The white people, especially the Belgians seriously mistreated them.

Appendix J. Reading Path Analysis. Mamie's Dream
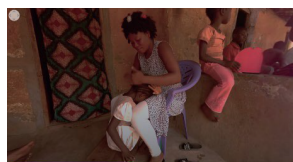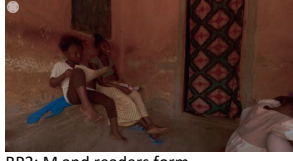
	1 IN BOAT	
	 perspective given at the onset	
	 lines up with Mamie in next shot	
	FTB: completion + spatiotemporal shift	
2 AT WATERSIDE		
	 SPF: Matched with previous view	
	 EPF1: When focus remains on M	 EPF2: When RP follows M gaze vector
Hard cut enhances action in next		
3_KIDS PLAYING FOOTBALL		
	 SPF1: Matches M in previous	 SPF2: lines up with river in previous

		 RP: as continued from SPF2 OR as directed from onlookers' collective gazes in SPF1	
 Follows strong kinetic vectors picked up from either SPF -RPs			
4_UNDER TREE IN YARD			
 SPF: Follows kinetic vector of ball in previous shot	Mamie if taken from asymmetric type comp (ball roles to left of frame)		
	 RP: camera movement leads focus		
	 RP: M becomes larger and salient throughout movement – have access to her gaze		
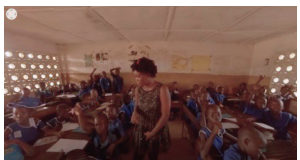
			 RP: M gaze (with head move) directs info focus to girls playing	
 EPF1: girls playing in slow motion - If target of M gaze is RP	 EPF2: If movement is primary RP motivation AND If M gaze seen as disengaged (resting in hands - no target)	 EPF3: if our view is centred on M and not applying an asymmetric composition		
5 STANDING IN DOORWAY				
Transition: cross dissolve spatio-temporal + compositional overlay with M represented twice (below)				
NEED THIS ONE	 Cross-dissolve effect (2 Mamies)			
	 SPF1: M emerging from dissolve with M in previous	 SPF2: taken from M centred in previous		
	 RP: M/bag framed (nested) by building and door			

 EPF1: Bike has salience. Equal in scale to M and proximity to M suggests transitive relevance	 EPF2: lines with Teacher in next			
Fade to black: signals spatiotemporal shift and completion				
6_#1 SCHOOLROOM				
Fade from black: signals new sequence of information				
 SPF1: continues from motorcycle	 SPF2: continues from M (repeats doorway)			
	 RP1: Emergent teacher enacts transitive sequence. Continues SPF2 Pupil gaze leads to teacher (although weak and not clearly directing focus)			
			 RP2: clear movement trajectory enacting clear ideational process – walking to and picking up	

		EPF: completion of RP and clear ideational process – clearing off (blackboard)		
7 SITTING ON BED				
Hard cut				
		SPF: follows teachers EPF.		
	RP: child's reaactive movement and gaze move focus to M + D.			
	EPF: Strong kinetic/transactional vectors keep focus here Dissolves to M in next shot			
8 IN MARKET				
	SPF: matches M+D in previous. Reinforced by clear framing/salience			

EPF1: matches M in next shot.		EPF2: shot lacks strong vectors. Some gaze vectors but not focused. Does not line up cleanly with M in next		
Fade to black				
9 SITTING UNDER SHELTER				
Fade from black				
	SPF1: M + D as salient figure	SPF2: follows M from previous Provides split view interior/exterior (liminal) M partially visible on left of view		
	RP1: follows either SPF1 or 2 to M as salient in int. area			
RP2: M and readers form compositional unit (inverted triangle through readers' feet) 2.a: Gaze points back to M				

	RP3: M's gaze directs focus out to open area Is unfocused and weak			
		 RP4: cooking woman incorporated in gaze (weak/indeterminate) Has salience via proximity/size/movement Gazes back (briefly) toward M		
	 EPF: M's diegetic audio (speaking playfully with child) draws attention back to her and her ongoing activities with daughter LINE WITH SPF1 IN NEXT			
Transition: cross-dissolve				
10_#2 SCHOOLROOM_1				
Transition: cross-dissolve				
			LINE WITH EPF IN PREV SPF1: look-room visible to left. Lines up with M in last shot and supports look space directionality > teacher	 SPF2: M centred
				SEEING WHO SHE IS WITH

7

		PR1: Follows M's focused gaze and orientation to > teacher (heard in diegetic sound) RP1a: Teacher gains salience as framed by blackboard and action sequence (writing, talking)		
		 RP2:Teacher engages verbally and via gesture and gaze with (audible) > classroom inhabitants		
	 RP2.a: Pupils audibly interact with teacher RP3: Brief (covert) focused gaze by pupils in M's seated area towards M's position			
 EPF2: M – if attentional 'loop' is completed with her via RP5	 if final point on path is pupils responding to teacher's gaze/verbal/actions			
11_#2 SCHOOLROOM_2				

 M1		 T1		
	 T2/M2			
				 EPF: Pupils oriented towards teacher's position react to teacher's prompt

12_#2 SCHOOLROOM_3

				 SPF1: Matches EPF in previous
			 RP1: centred in view, extends hand and points finger towards pupils	

		RP2: Pupils emphatically respond to M with gaze and body movements (raised hands) and verbally		
		<< >> RP3: Narrow angle between participants; RP follows ongoing bilateral exchange between M and PUPILS		
			EPF:	
13 OPEN AREA				
			SPF: M matches M in previous. Salient as moves synchronously with camera	
			RP1: Acknowledges and gesticulates towards cluster of seated people >>	RP1a: Little if any response from cluster – camera moves away with M continuing her salience

		 RP: Stops to speak with / gesticulate towards cooking woman		
			 RP: moves towards cluster of seated/standing people – raises hand in gesture/greeting	
			 EPF: M remains salient figure to end. Matches SPF in next with M salient in triangular composition	
14 IN CHURCH				
			 SPF: Triangular composition with M salient	
RP: no clear vectors or compositional paths to follow from SPF. Shot inhabitants are internally focused (singing/moving). Homogenous field of activity to be scanned				
			 EPF: lines up with M and girl in next shot	

			MATCH WITH SPF	
15 HOLDING HANDS				
			SPF: Mamie and girl. Matching M in previous shot	
				EPF: M+G. Matches Girl in next shot
16 GIRLS PLAYING				
		SPF: problematic. Matches previous EPF but figure (girl) has little/no salience or ideational continuity. Is RP (option 1 – scanning)		
			RP(op2)1: One girls gaze directs attention to an unseen figure (male adult)	

				RP(op2)2: Male adult stands beside salient white structure – gesturing and looking back towards dancing girls
			RP(op2)3: Dancing girls increase efforts at dance	
			EPF: Matches M and girl in next shot	
17 SERIOUS CHAT				
			SPF: M and girl at personal distance framed by doorway	
			EPF: no suggested RP away from M and girl	
18 DANCING CLAPPING				
			SPF: Mamie matches previous EPF. M is highly salient at personal distance and kinetic	

			RP1: Mamie dancing and clapping. Gaze is upward and outwards to crowd (disengaged); downwards to crowd (engaged)	
				 RP2: Crowd is oriented and gazing towards stage area, bringing attention back to stage area (Mamie
			 EPF: target of crowd gaze. Matches movement vector in next shot	
19 RIVER				
			 RP1: M becomes larger and more salient via temporary proximity to camera and as sole figure in shot	
		 RP2: M's temporary salience allows for reverse view of M front		

		RP3: In continued view of M front, M becomes small relative to surrounds and begins to merge with the environment		