Multimodal System for Public Speaking with Real Time Feedback: A Positive Computing Perspective

Fiona Dermody
Dublin City University
Dublin, Ireland
Fiona.Dermody3@mail.dcu.ie

Alistair Sutherland
Dublin City University
Dublin, Ireland
Alistair.Sutherland@dcu.ie

ABSTRACT
A multimodal system for public speaking with real time feedback has been developed using the Microsoft Kinect. The system has been developed within the paradigm of positive computing which focuses on designing for user wellbeing. The system detects body pose, facial expressions and voice. Visual feedback is displayed to users on their speaking performance in real time. Users can view statistics on their utilisation of speaking modalities. The system also has a mentor avatar which appears alongside the user avatar to facilitate user training. Autocue mode allows a user to practice with set text from a chosen speech.

CCS Concepts
•Human-centered computing□Interface design prototyping
•Human-centered computing□User interface design

Keywords
Multimodal Interface; Affective Computing; Human Computer Interaction; Positive Computing; Public Speaking

1. INTRODUCTION
Real-time feedback has been utilised in multimodal systems for public speaking training such as [2], [4], [5], among others. In previous work, our system displayed textual and visual real-time feedback on all speaking modalities on the periphery of the screen [3]. In this demonstration we present an updated system with intuitive visual feedback located in proximity to the avatar, a new mentor avatar and autocue mode.

2. OVERVIEW OF SYSTEM
We present a positive computing system for giving feedback in real-time to a user practicing to make a speech in a public-speaking context. The system uses the Microsoft Kinect to sense the user's body movements, facial expressions and voice. Feedback is given on a laptop screen in front of the user. The system has been developed within the framework of positive computing [1]. The objective of the system is to reduce the anxiety which a user experiences when speaking in public. From this objective, it follows that the system itself must not add to any anxiety already experienced by a user. Using the system should be an enjoyable experience. Furthermore, users should feel that using the system is beneficial for them in addressing any speaking anxiety. Feedback displayed by the system should be assimilated with minimum cognitive load on the users. If users are stressed trying to assimilate feedback from the system, they are unlikely to either react to it or find the system pleasant to use.

We have focused on the following tenets of positive computing – autonomy, self-awareness and competence. In the context of this system, autonomy relates to the user's ability to choose what feedback to focus on. Self-awareness relates to the user gaining an insight into how they appear while speaking through observing the system avatar. Competence refers to the user's ability to utilize different speaking modalities such as eye contact, gestures and vocal variety. Initial user reaction to the system has been positive with regards to reported autonomy, self-awareness and competence.

3. SYSTEM COMPONENTS
3.1 Feedback in Real Time
Our previous system provided continuous feedback in real time [3]. The system had an avatar, which imitated the user's body movements and mouth and eyebrow movements. Feedback was given by text and icons around the periphery of the screen. This meant that it was challenging for users to attend to the avatar and assimilate the feedback at the same time. The updated system presented in this paper has a similar avatar but displays feedback using visual icons displayed in proximity to the avatar, Figure 1. These icons include the following:

1. Arrows around the avatar's head to prompt the user to change their view direction.
2. A rolling graph which displays the pitch of the user's voice. This allows the user to see how fast they are talking and to gauge whether they are talking in a monotone voice or using a lot of vocal variety.
3. An icon to indicate when the user's hands are touching. The icon is located over the avatar's hands
4. An icon to indicate when the user has crossed their arms
5. An icon to indicate if the user is agitated or moving too quickly. The icon is located next to the avatar's body
6. An icon to indicate whether the user is smiling or surprised. The icon is located next to the avatar's face.

ICMI'16, November 12–16, 2016, Tokyo, Japan
ACM. 978-1-4503-4556-9/16/11...$15.00
http://dx.doi.org/10.1145/2993148.2998536

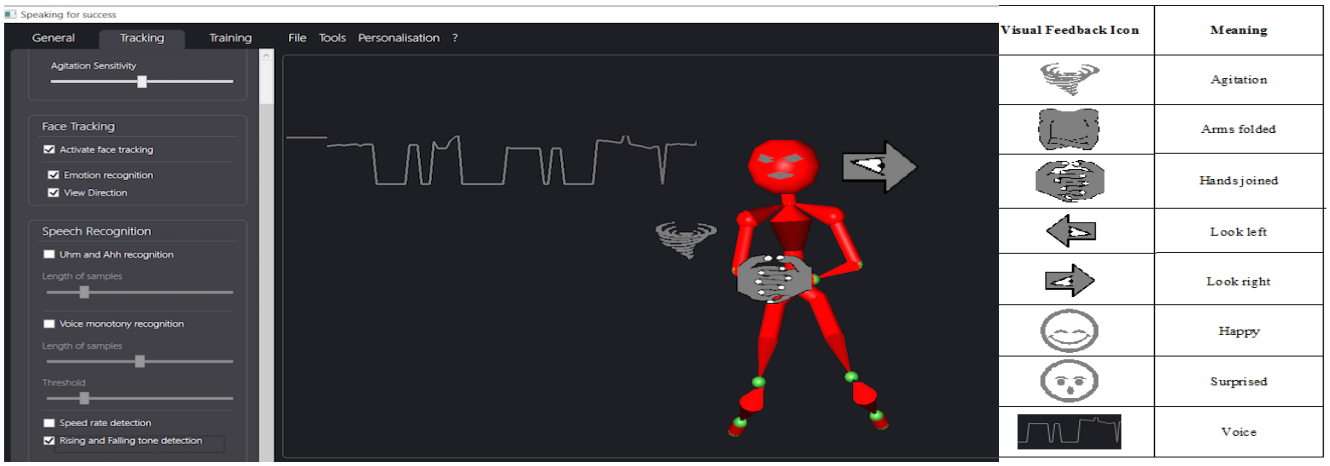

Figure 1: (a) Indicative visual feedback – displayed in proximity to avatar (b) Visual feedback icons

3.2 Charts and History

The system records statistics about the users performance over a period of time. These can be displayed as charts so that the user can monitor their progress. For a particular practice session, the user can see a timeline, which displays when they made certain actions e.g. crossed their arms, touched their hands, became agitated etc. They can also display histograms of their progress over days or weeks. This allows the user to see if they make more gestures or move more than they did in the past.

3.3 Avatar Training

The system displays a "mentor" avatar, which appears in the same space as the user's avatar, see Figure 2. The mentor avatar can illustrate how a certain action should be performed e.g. it may give an example of a gesture. The user can then imitate the action of the mentor and then the mentor will give feedback on the user's performance. For example, it may say "Perform the action more quickly" or "Place your hand higher". It could then perform the action again to indicate how the user should modify their action.

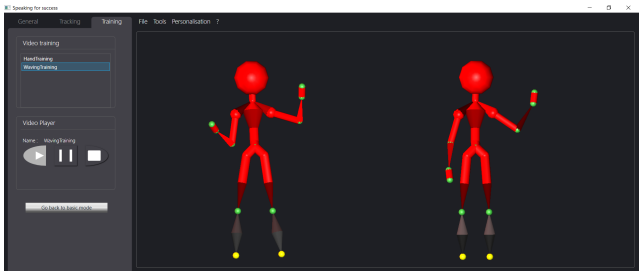

Figure 2: Avatar Training – Mentor Avatar on Left and User Avatar on Right

3.4 Voice Recognition and Autocue Mode

The system now has improved voice recognition. If the user is giving a set speech, where the text is known in advance, the system can follow the users words and prompt them with the next line in the speech.

4. CONCLUSION

We have developed a multimodal system for public speaking with real-time feedback within the framework of positive computing using the Microsoft Kinect. Real-time visual feedback is displayed to users on their speaking performance. Users can view statistics on their utilisation of speaking modalities. The system also has a training avatar and autocue mode.

5. ACKNOWLEDGMENTS

This material is based upon works supported by Dublin City University under the Daniel O'Hare Research Scholarship scheme. System prototypes were developed in collaboration with interns from École Polytechnique de l'Université Paris-Sud and l'École Supérieure d'Informatique, Électronique, Automatique (ESIEA) France.

6. REFERENCES

[1] Calvo, R.A. and Peters, D. 2014. Positive Computing: Technology for wellbeing and human potential. MIT Press.

[2] Chollet, M. et al. 2015. Public Speaking Training with a Multimodal Interactive Virtual Audience Framework - Demonstration. ICMI 2015 - Proceedings of the 2015 ACM International Conference on Multimodal Interaction. (2015), 367–368.

[3] Dermody, F. and Sutherland, A. 2015. A Multimodal System for Public Speaking with Real Time Feedback. Proceedings of the 2015 ACM International Conference on Multimodal Interaction (2015), 369–370.

[4] Hoque, M.E. and Picard, R.W. 2014. Rich Nonverbal Sensing Technology for Automated Social Skills Training. Computer. 47, 4 (2014), 28–35.

[5] Schneider, J. et al. 2015. Presentation Trainer, your Public Speaking Multimodal Coach. Proceedings of the 2015 ACM on International Conference on Multimodal Interaction (2015), 539–546.