



# A Multi-Modal System for Public Speaking

<sup>1</sup>Fiona Dermody <sup>2</sup>Dr. Alistair Sutherland <sup>3</sup>Dr. Margaret Farren  
<sup>2,3</sup>(Supervisors)

<sup>1,3</sup>Faculty of Humanities

<sup>1,2</sup>Faculty of Engineering and Computing  
Dublin City University, Dublin 9. Ireland.

Fiona.Dermody3@mail.dcu.ie

**Abstract.** This research involves the development of a digital system and user interface to analyze social signals displayed during public speaking and to provide real-time feedback to users on speaking performance.

**Keywords:** Human Computer Interaction, Affective Computing, Multi-Modal Interfaces

## 1 Research Topic Details

### 1.1 Description of Research Problem to be solved

Can a digital system be developed to recognize the multi-modal social signals of effective public speaking and through the provision of feedback enable users to become more effective at public speaking?

### 1.2 Why it is important to solve this research problem

The fear of public speaking tops the list of human phobias. However, success in social, academic and occupational situations depends on the ability to communicate effectively to groups. A fear of public speaking thus limits achievement in social gatherings, education and enterprise [1]. However, the fear of public speaking can be so great that it can lead to avoidance of speaking in the public domain altogether [2]. The problem is a recursive one, how can an individual improve their speaking skills and reduce their fear of speaking in public if they avoid speaking in public?

Public speaking is not just about the words spoken. Effective speaking involves the use of gestures, facial expressions and vocal variety. All these

social signals combine to give the appearance of self-confidence in a speaker. Research has found that anxious speakers do not engage the attention of an audience [3]. Rather the audience focuses on the speaker's nervous disposition instead of their words [4].

The solution, which we propose, is interdisciplinary. It incorporates theory from computing, psychology and communications. It is envisaged that this social signal recognition system will enable individuals to develop their competence in public speaking.

Using a combination of 3D video imaging, audio and social signal processing algorithms, this digital system analyses facial expressions, tone of voice and gestures. The system then provides feedback on the user's speaking performance. It will also deliver tutorial videos on good speaking practices. Exposure to these dynamic features will enable a speaker to systematically develop confidence and skill before speaking in front of a live audience.

One of the technical challenges to be overcome during the development of the system is the optimal way to display feedback to the user in real time during their speaking task. The nature of the feedback is imperative, as our survey showed that some users prefer visual feedback while others prefer textual feedback. We decided to allow feedback to be customizable on the interface.

Skilled human trainers in communication are scarce and expensive. This digital system incorporates experience from one such skilled human trainer to provide constructive feedback to users on their speaking performance. This digital experience will enable anxious speakers to develop their public speaking skills cost-effectively, in private and at their own pace.

By harnessing the power of social signal processing, this system will increase the user's communication skills, confidence and ultimately lead to greater success in life.

### 1.3 Justification that prior research has not solved this problem

There have been attempts to use social signal recognition for public speaking but not for instruction purposes [5]. This system is innovative because it will extend the field of human computer interaction:

- Combining all modalities - voice, gesture, facial expression and body pose into a multi-modal system for delivering instruction in public speaking
- Incorporating the knowledge of experts in public speaking ensures that the feedback provided on a user's speaking is based on a real-world, practice-based approach

## 2 Research Hypotheses

- A multi-modal system can accurately recognise the characteristics of gestures, body language, voice, facial expression that influence an audience during public speaking
- Real-time multi-modal feedback is most effective for users to develop skill in public speaking

## 3 Methods used to develop and evaluate system

### 3.1 Prototype Development

A prototype has been developed following a user survey on the features required in a multimodal system for public speaking.

### 3.2 Evaluating the work and presenting credible evidence of results to the research community

- Using the knowledge of experts in public speaking to construct a tutorial system, which will give automatic feedback
- Carrying out controlled experiments on volunteer users to measure the efficacy of the system.

## 4 Sketch of the proposed solution

The system will use a Microsoft Kinect connected to a computer. Social Signal Processing techniques will be used to recognize the speaker's body language, gestures, voice, and facial expressions

- Classify speaker's emotion, as perceived by the audience, from the combination of the above
- Perceived emotion is regarded as the primary component for analysis.
- Give feedback to user on speaking performance
- Provide examples of good and bad speaking practice
- Set tutorial exercises from beginners level to advanced level and will evaluate the user's performance

## 5 Expected contributions of the PhD Research

There have been attempts to use social signal recognition for public speaking but not for instruction [5]. Our research will extend the field of human computer interaction -

- To combine all modalities (voice, gesture, face, etc.) into a multi-modal system for public speaking
- To construct models for different types of speech, speaking styles, contexts
- To use the knowledge of experts in public speaking to construct a tutorial system, which will give automatic feedback
- To carry out controlled experiments on volunteer users to measure the efficacy of the system.

## 6 Statement of work to date

- User survey on features required in multimodal system for public speaking
- A prototype has been developed
- Initial user testing has started

## 7 Issues for discussion at the Doctoral Consortium

- Prototype interface will be presented for discussion
- Is prototype interface design effective?
- Is the feedback presented in a clear and effective way?

## Acknowledgements

This material is based upon works supported by Dublin City University under the Daniel O'Hare Research Scholarship scheme.

## References

- [1] Dwyer, K. K., & Davidson, M. M. Is Public Speaking Really More Feared Than Death? *Communication Research Reports*, 29(2), 2012, pp. 99-107.
- [2] Feldman, P. J., Cohen, S., Hamrick, N., & Lepore, S. J. Psychological stress, appraisal, emotion and cardiovascular response in a public speaking task. *Psychology & Health*, 19(3), 2004, pp. 353-368.
- [3] Freeman, T., Sawyer, C. R., & Behnke, R. R. Behavioral inhibition and the attribution of public speaking state anxiety. *Communication Education*, 46(3), 1997, pp. 175-187.

- [4] Winton, E. C., Clark, D. M., & Edelmann, R. J. Social anxiety, fear of negative evaluation and the detection of negative emotion in others. *Behaviour Research and Therapy*, 33(2), 1995, pp. 193-196.
- [5] Batrinca, L., Stratou, G., Shapiro, A., Morency, L. P., & Scherer, S. Cicero-towards a multimodal virtual audience platform for public speaking training. In *Intelligent Virtual Agents*, Springer Berlin Heidelberg, 2013, pp. 116-128.