[Poster] Exploring Social Augmentation Concepts for Public Speaking using Peripheral Feedback and Real-Time Behavior Analysis

Ionut Damian* Chiew Seng Sean Tan† Tobias Baur* Johannes Schöning† Kris Luyten†
Elisabeth André*

Abstract

Non-verbal and unconscious behavior plays an important role for efficient human-to-human communication but are often undervalued when training people to become better communicators. This is particularly true for public speakers who need not only behave according to a social etiquette but do so while generating enthusiasm and interest for dozens if not hundreds of other persons. In this paper we propose the concept of social augmentation using wearable computing with the goal of giving users the ability to continuously monitor their performance as a communicator. To this end we explore interaction modalities and feedback mechanisms which would lend themselves to this task.

1 Introduction

Inter-human communication is a complex multi-layered activity which, although we partake in it on a regular basis, is very difficult to understand and master. In particular, nonverbal communication, which is known to play a larger role during an interaction than its verbal counterpart [9, 8, 3], is very challenging to control during interaction as it is mostly unconscious. Birdwhistell argued that communication control is not achieved through an associative process which involve explicit mapping of words and gestures during sentence construction, but rather through perceived patterns that have been learned [3]. For example, we rarely think about what gestures or postures we perform while interacting with other persons. Thus, controlling the information flow generated by the non-verbal behavior in an interaction is a challenging task, especially in the context of public speaking.

Most of the current practices for public speaking training involve either a priori learning (e.g. reading books, rehearsals) or a posteriori feedback (e.g. performance analysis[1]). Technology-enhanced feedback during the actual interaction is still virtually unexplored, as it involves an additional information flow which, in most cases, would interfere with the actual interaction.

In this paper we propose to address this issue by making use of modern wearable devices to achieve behavior monitoring with the ultimate goal of enhancing the user's awareness of their own nonverbal behavior during human-human interactions, such as public speaking, as well as providing guidance towards improving it. We call this concept social augmentation, as it enhances awareness of the user's own behavior and guides her towards improving it. Over the past few years, we witnessed a boom in wearable devices, including head mounted displays (HMD) such as the Google Glass. These devices enable users to perceive a myriad of information with minimal effort, thus, allowing the users to carry out normal activities at the same time. This makes such devices ideal for our task. More specifically, using Augmented Reality (AR) concepts, users can receive information about their nonverbal behavior during stressful social situations which take the user out of their comfort

*Augsburg University, Germany
†Hasselt University - tUL - iMinds, Belgium

Figure 1: Prototype system using optical HMD (left) and illustration of the peripheral feedback from the user's point of view (right).

zone, for example, going on a date, giving a lecture to an audience, or going to a job interview. Furthermore, the wearable nature of HMDs means the system is not location-bound nor context-specific and does not require setting up before the interaction. In these examples, the system could help the user make a better impression on the interlocutors. More precisely, in a public speaking scenario, the system can make the user aware if the speech is boring or she talks too fast.

2 Social Behavior Monitoring and Interpretation

To be able to offer the user both an analysis of her social behavior as well as feedback regarding how to improve it, we first need to capture the low level signals from various sensors and transform them to high level features such as gestures and facial expressions. This can be achieved with a social signal processing framework. However, what behaviors to analyze and which sensors to use is scenario specific. For example, in a public speaking scenario, where interaction is limited to a single room, remote sensors such as user facing cameras can be used. This is not the case when the interaction location is dynamic, e.g. a dating scenario. Furthermore, other factors such as intrusion, accuracy, size and privacy issues must also be considered. The timing of the behavior analysis is also a crucial aspect of the system. If the feedback is delivered too late it might become obsolete, thus confusing the user. To this end, it is paramount that the social signal processing happens in real-time with minimal recognition delay.

Considering our public speaking scenario and given the current state of the art, the following behaviors are of relevance and feasible to use: gestures, postures, facial expressions, vocal quality (pitch, speechrate, loudness) and expressivity (movement energy, spatial extent, fluidity)[2, 7]. This selection would require a user-facing depth camera (or user-worn motion capturing system) for gestures, postures and expressivity features [2, 5], a user-facing video camera (or facial physiological sensors) for facial expressions [2, 6] and a microphone for audio analysis [2].

3 Peripheral Feedback

In order to communicate the behavioral analysis to the user, we propose a peripheral feedback approach where functional icons are displayed in the periphery of the user's field of view using an optical see-through HMD. The peripheral feedback is designed to increase the self-awareness on one's social behavior and to offer guidance regarding its improvement. In this context, the design of the icons

Figure 2: A sample of the functional icon design space for three behaviors at different intensities.

is a critical aspect as it defines the usability of the whole system and it may have an effect on the user's social interaction. For instance, mutual gaze is one of most essential things in social communication. There is the danger that the peripheral feedback distracts the user from the actual interaction, blinding her against the behaviors of the interlocutors. Additionally, the user might be perceived as not keeping eye contact if the peripheral feedback draws too much focus. On the other hand, breaking eye contact from time to time is normal in social interactions as uninterrupted eye contact can be perceived as staring. To address such problems, we defined a set of design guidelines based on Glassware development design principles[1] which we adapted to our task:

- **Feedback Alignment**. The peripheral feedback should be provided in the upper left or right corner of the user's field of view, above the natural line of sight. This allows users to engage in an interaction with others and choose for themselves when to access the feedback. Furthermore, a near focal plane should be chosen to further discriminate the behavior analysis from the scene as the feedback is intended to address the user's own (and private) behavior.
- **Feedback Presentation**. Only the most relevant behavioral feedback should be visible to the user. The icon design should be pictographic, flat and not too detailed to avoid cluttering and make them easy to understand. Furthermore, the icons should not rely solely on color to encode information as color retrieval is susceptible to background and lighting conditions when using see-through HMDs.
- **Persistent and continuously updated status**. The peripheral feedback should not distract the user by popping in and out of the user's field of view at unexpected times. Hence the feedback should be designed to be always-on and persistently show the behavioral analysis, thus, giving the user the ability to choose when to access the information.
- **Ephemeral status access**. The feedback should be designed to focus on ephemeral access to the behavioral analysis, which is akin to a fire-and-forget usage model. In this manner, the users can glimpse at the feedback and quickly extract the behavioral information without interrupting their main activity.

Using these guidelines, we defined a design space for functional icons depicting various behaviors. An extract of this design space can be seen in Figure 2.

4 System Prototype

To test out the concept, we developed a first prototype using a see-through Vuzix Star 1200 for the peripheral feedback (Figure 1). The peripheral feedback is displayed monoscopically and consists of three icons which are continuously displayed in the upper right corner of the HMD's right display. Each icon can show three levels of intensity (low, medium, high) based on the behavioral analysis. Additionally, to offer the user information regarding the appropriateness of her behavior, each icon is accompanied by either a small green ✓ (in case the behavior is appropriate) or a red ✗ (if it is not).

For this prototype we focused on analyzing a subset of the behavioral cues proposed in Section 2. To this end we employ the Social Signal Interpretation (SSI) framework [11] using a depth camera and a microphone. This way we are able to analyze the speech rate, movement energy and excitement level of the user in real-time.

Speech rate is computed from the voice of the user using a PRAAT analysis [4] whereas the movement energy is computed form the position of the user's hands over time. In order to compute the excitement value, we first infer the user's affective state from the postures she displays in a certain time window [10]. From the affective state we extract the value along the boredom-excitement dimension which gives us the excitement level of the user. For all behaviors, a thresholding approach is applied to convert the raw values to the three intensities the icons are able to display.

5 Discussion and Future Work

The use of an HMD for social augmentation might result in some problems when used in the real world which the designers must be aware of. Section 3 addresses some of this issues, however a full scale user study is required to fully understand the implications of this concept. Privacy is another concern which needs to be considered, especially if the system's sensors can also capture the behavior of other persons.

Overall, in this paper we explored the concept of using social signal processing and peripheral feedback to augment user's ability to sense and control their social behavior. For our future work we will further build upon the design space of the functional icons. Furthermore, we intend to evaluate the system in a real scenario in order to determine the usability and feasibility of the whole system as well as the impact it has on the user's social skills.

[1] https://developers.google.com/glass/design

References

[1] L. M. Batrinca, G. Stratou, A. Shapiro, L.-P. Morency, and S. Scherer. Cicero - towards a multimodal virtual audience platform for public speaking training. In *IVA*, volume 8108 of *LNCS*, pages 116–128. Springer, 2013.
[2] T. Baur, I. Damian, P. Gebhard, K. Porayska-Pomsta, and E. André. A job interview simulation: Social cue-based interaction with a virtual character. In *Proc. SocialCom 2013*, pages 220–227. IEEE, 2013.
[3] R. L. Birdwhistell. *Kinesics and context: Essays on body motion communication*. University of Pennsylvania press, 2011.
[4] P. Boersma and D. Weenink. Praat: doing phonetics by computer [computer program]. version 5.3.56, 2013.
[5] I. Damian, M. Obaid, F. Kistler, and E. André. Augmented reality using a 3d motion capturing suit. In *AH '13*, pages 233–234. ACM, 2013.
[6] K. Fukumoto, T. Terada, and M. Tsukamoto. A smile/laughter recognition mechanism for smile-based life logging. In *Proc. AH 2013*, pages 213–220, New York, NY, USA, 2013. ACM.
[7] P. Gebhard, T. Baur, I. Damian, G. Mehlmann, J. Wagner, and E. André. Exploring interaction strategies for virtual characters to induce stress in simulated job interviews. In *Proc. AAMAS 2014*, pages 661–668, 2014.
[8] K. Hogan. *Can't Get Through: Eight Barriers to Communication*. Pelican Publishing, 2003.
[9] A. Mehrabian. *Silent messages: Implicit Communication of Emotions and Attitudes*. Wadsworth Publishing Co Inc, Belmont, 1981.
[10] C. S. S. Tan, J. Schöning, K. Luyten, and K. Coninx. Informing intelligent user interfaces by inferring affective states from body postures in ubiquitous computing environments. In *IUI*, pages 235–246, 2013.
[11] J. Wagner, F. Lingenfelser, T. Baur, I. Damian, F. Kistler, and E. André. The social signal interpretation (SSI) framework - multimodal signal processing and recognition in real-time. In *Proc. ACMMM 2013*, 2013.