Augmenting Social Interactions: Realtime Behavioural Feedback using Social Signal Processing Techniques

Ionut Damian*, **Chiew Seng Sean Tan**‡, **Tobias Baur***, **Johannes Schöning**‡,
Kris Luyten‡, **Elisabeth André***
*Human Centered Multimedia, Augsburg University, Germany
‡Expertise Center for Digital Media, Hasselt University - tUL - iMinds, Belgium
{damian, baur, andre}@hcm-lab.de, {sean.tan, johannes.schoening, kris.luyten}@uhasselt.be

Figure 1. Concept of the *Logue* system: A user wearing an HMD while giving a public speech (left). Using various sensors and social signal processing techniques (not illustrated), the user receives realtime feedback on his own behaviour superimposed on field of view (right).

ABSTRACT
Nonverbal and unconscious behaviour is an important component of daily human-human interaction. This is especially true in situations such as public speaking, job interviews or information sensitive conversations, where researchers have shown that an increased awareness of one's behaviour can improve the outcome of the interaction. With wearable technology, such as Google Glass, we now have the opportunity to augment social interactions and provide realtime feedback on one's behaviour in an unobtrusive way. In this paper we present *Logue*, a system that provides realtime feedback on the presenters' openness, body energy and speech rate during public speaking. The system analyses the user's nonverbal behaviour using social signal processing techniques and gives visual feedback on a head-mounted display. We conducted two user studies with a staged and a real presentation scenario which yielded that *Logue*'s feedback was perceived helpful and had a positive impact on the speaker's performance.

Author Keywords
Computer-Enhanced Interaction; Behaviour Analysis; Peripheral Feedback; Social Signal Processing

ACM 978-1-4503-3145-6
http://dx.doi.org/10.1145/2702123.2702314

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

INTRODUCTION & MOTIVATION
When communicating with others we tend to use several modalities simultaneously. The most obvious way of communication is speaking to each other (verbal communication). However, our body also plays a very important role during direct communication. It communicates subtleties and emotions to others related to what we are talking about and in which context we are addressing others. In the last decades other researchers have shown that nonverbal behaviours, such as gestures, facial expressions or the way we use our voice, play a more significant role during an interaction than its verbal counterpart [5, 17, 26]. To become an efficient communicator, mastering appropriate usage of both verbal and nonverbal communication is essential.

For particular situations such as a public speech, a sales conversation or a job interview, people often need to train and practice both their verbal and nonverbal communication skills, because they can substantially contribute to reaching the set targets, e.g. getting hired [9, 12, 18]. Traditionally, training is done off-line in practice sessions with experts that provide feedback. One well-known approach is using role-plays with a coach that are recorded on video, with the recordings being analysed and discussed afterwards.

We present *Logue*, a system designed to provide in-situ and realtime feedback on a speaker's nonverbal communication unobtrusively during a public speaking scenario using a wearable display (Figure 1). The feedback aims to increase the

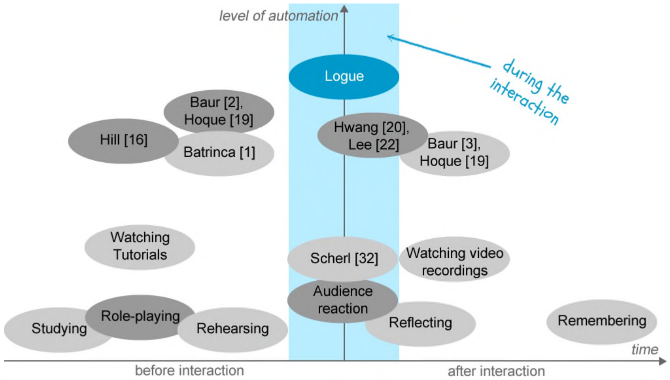

Figure 2. *Logue* put in relation to current training practices and related work.

users' awareness of their own nonverbal behaviour, as well as inform of the behaviour's appropriateness in the given scenario. *Logue* employs social signal processing techniques to analyse the speaker's performance using data from a microphone and a depth camera. Based on this processing, we can then inform the user on the speech rate, body energy and openness in realtime using an HMD.

We also discuss the challenges that arise from augmenting social interactions and use these as drivers for the design of our system. Specifically, we focus on how to provide the users with an additional information channel (behavioural feedback) without distracting them from their main task. Over the course of two user studies, we evaluate the effect of *Logue* on the user's behaviours as it is measured (A) by the system using signal processing, (B) by other persons and also (C) by the users themselves using surveying techniques. We show that *Logue* has a positive effect on the performance of the users, enabling them to adapt their behaviour in realtime.

RELATED WORK

Progress made in the area of social signal processing has enabled the development of novel tools for training social skills. These tools record and automatically analyse social and emotional signals while human learners engage in a social interaction. Popular scenarios for training social and emotional skills include job interviews [2, 19], public speeches [1, 23], military scenarios [16] or for social anxiety [28] or stress training [24]. In many cases, the analysis of social and emotional signals is done in realtime, for example, to enable the virtual character that simulates a human interlocutor to immediately respond to it. However, the human learner is usually not given any explicit feedback on her or his performance before the end of the interaction. For instance, the TARDIS system [3] enables the learners to inspect previous interactions conducted with a virtual character serving as a recruiter in a job interview. In this paper, we present an approach that provides realtime feedback on social behaviours: directives are automatically generated by a computer system while the user is engaging in a social interaction.

Live feedback has already been successfully employed in psychotherapeutic training sessions. Various devices, such as ear phones and teleprompters, have been explored to present conversational aids selected by a human supervisor. An example of such a training scenario is presented by Scherl and Haley [32]. While a trainee therapist is talking to a patient, he or she is provided with directives from a human supervisor that are displayed on a computer monitor. An evaluation revealed that computer-supported live feedback can be provided effectively during conversations between trainee therapists and their patients as long as the directives from the human supervisor are clearly formulated and kept short. However, the conversational aids are manually produced by a human coach. In contrast, our work presents an approach that generates directives automatically, drawing on the concept of augmented social interaction.

Figure 2 places our proposed concept in relation to known training practices and related work. The horizontal axis describes the timing of the training relative to the interaction itself whereas the vertical axis categorizes the methods based on their level of automation.

AUGMENTING SOCIAL INTERACTIONS

Starting from Engelbart's framework for augmenting human intellect [13], Xia and Maes [38] identify new cognitive domains for personal augmentation including memory, motivation, decision making, and mood. In our work, we propose social augmentation as a novel domain based on the consideration that social behaviour is at the core of human intellect. Social augmentation includes the use of physical items, such as miniaturized sensors and light-weight displays, to provide users with automatically generated realtime feedback on their nonverbal behaviours.

Earlier studies by Scherl and Haley [32] have shown that the display of conversational aids during social interactions may improve communication. Furthermore, studies by Ofek and colleagues [27] seem to indicate that secondary information can be consumed by users during a conversation without the interlocutors noticing it. Despite encouraging findings, there are also qualified concerns that social communication might be impaired by secondary information that interferes with it.

A particular challenge is the dual-purpose use of modalities. On the one hand, gestures may be employed to control the social augmentation interface. On the other hand, they may serve to accompany speech in social interactions. To avoid disruptions resulting from the dual-purpose use of modalities, interactions to control the interface have to be kept at a minimum or even avoided completely.

Similarly, eye gaze is also susceptible to problems regarding dual-purpose use of modalities. More precisely, there is a risk that wearable displays lead to reduced gaze contact because the user needs to split her or his attention between the display and the social interlocutor. Studies by McAtameny and Parker [25] point out some of the issues arising from the use of a head-mounted display during social interactions. While wearing a non-activated head-mounted display does not seem to influence how others perceive the user, an activated display may have a negative impact on the social interaction.

To solve these issues, a number of techniques have been developed to provide secondary information in a way that does not disrupt social interaction. Ofek and colleagues [27] conducted various experiments to derive design guidelines for the presentation of secondary information on a head-up display. They found that users can process information better when it's delivered in batches rather than sequential. Furthermore, they strongly advise against the use of auditory cues while the user is speaking. Experiments by deVaul and colleagues [11] demonstrated the potential of subliminal cues as a way to present memory aids in a non-distracting manner.

Our paper extends previous research by combing techniques for social signal processing with techniques for personal augmentation in order to provide people with live visual feedback on their social behaviour. While users are engaged in a social interaction with others, social communication aids are projected as overlays in their field of view. There are a number of options that might be taken into account to provide sociometric feedback. We excluded audio feedback because it might disturb users if they are producing sounds themselves (e.g. speaking), as noted by [36, 27]. Vibrotactile feedback has been successfully employed for teaching sensory-motor skills, for example, when playing a musical instrument [36]. However, for our purpose, haptic feedback seems less appropriate since speakers might have problems to relate a larger variety of haptic signals to specific conversational features. A number of systems that aimed to give feedback on communicative behaviours made use of mobile phones, for example, to support communication in face-to-face conversation [22], group meetings [21], during public speeches [31] or to help children with language delay [20]. Social augmentation techniques have also been explored in video conferencing settings [35] using normal desktop monitors. We decided to make use of head-mounted displays which bear the advantage that they do not require speakers to switch back and forth between the device and the interlocutor(s). Augmentation cues can be proactively provided to the users without requiring them to deliberately focus on the device. Furthermore, such devices are more suitable to provide personal information, such as feedback on the user's presentation behaviour, than external displays. External monitors might cause privacy problems as the feedback would be viewable by other persons as well. HMD's also allow the user to receive feedback while still looking in the direction of the interlocutor, which is of particular importance during social interactions. In this manner, we hope to keep the disruption of the social interaction as low as possible while giving guidance to users.

SCENARIOS

In this section we will present the reader with three concrete scenarios where social augmentation is feasible and discuss how it can benefit the users.

Job Interview

Job interviews, as a type of human-human interaction, rely on the ability of the interlocutors to read each other's behaviour and emotion. For the recruiter, the goal is to determine whether the interviewee is adequate for a specific job [30]. In this scenario, it is expected from the interviewee to remain cool and composed and not show extreme emotions [33]. More precisely, behaviours such as body expressivity (e.g. gestures, postures or facial expressions), vocal quality (e.g. speech rate, loudness, etc.) and eye gaze behaviour have been found to be problematic during job interviews [18, 8]. Considering this, the interviewee's nonverbal behaviour is a crucial element in the outcome of the whole interview and an increased awareness would only benefit this outcome.

The main challenge of this scenario is related to sensor choice. Considering the fact that a user attends job interviews in different locations, there is need for a fully wearable system that also ensures the full privacy of the applicant. To this end, sensors such as clothing embedded IMUs (inertial measurement units) or eye tracking glasses are preferable to remote sensing devices.

Information-sensitive conversation

During information-sensitive conversations, such as physician-patient conversations, ensuring sufficient behavioural awareness can be advantageous to the interaction goals. Blanch et al. [6] showed that the quality of nonverbal skills impacts how the physician is perceived by the patient. If we take the physician-patient conversation as an example, interaction augmentation can assist the physician with the delivery of sensitive information. For example, the system could remind the physician to use pauses to allow the patient to process the information, or provide feedback to increase vocal quality (e.g. speech rate, loudness) to boost the likelihood of the patient correctly understanding the message. Furthermore, an appropriate display of facial expression can also help avoid misunderstandings.

In order to allow the physician to be mobile and have consultations in different locations, body worn sensors are more appropriate than remote sensors. For instance, a microphone could be used for the speech and turn-related features and a head worn sensor, for example attached to a glass frame (or HMD), would allow the analysis of facial expressions [14].

Public Speaking

Public speakers need not only deliver a convincing message to their audience, but also inspire and generate enthusiasm at the same time. Thus, public speaking is a particularly stressful situation as the speakers need to master both their verbal and especially their nonverbal behaviour. For example, even a highly interesting message, if delivered with a reduced speech rate may cause boredom. Similarly, if the speech rate is too high, people might have difficulty understanding the message at all. The key lies in the balance.

In light of this, providing the user with feedback on nonverbal behaviour may help maintain this behavioural balance. For instance, feedback on speech rate and body energy would help the user to better convey excitement and enthusiasm, whereas increasing awareness on posture and gesture display can make the speaker appear friendlier and more approachable [29].

Unlike the other scenarios, remote sensors (e.g. video camera, depth camera, room microphone) are feasible for public

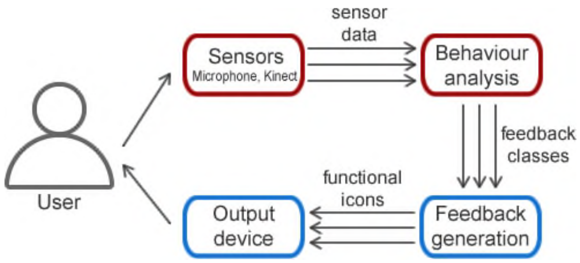

Figure 3. Structure of *Logue* showing the two main components, behaviour analysis and feedback generation, and the information loop.

speaking and do not suffer from the increased intrusion level associated with body worn sensors.

THE LOGUE APPLICATION
In order to get a better understanding of the feasibility of augmenting social interaction, we implemented the *Logue* system for the scenario of public speaking. *Logue* augments public speakers by providing them with realtime feedback on their nonverbal behaviour. The aim of the system is to increase the user's awareness of her or his own body language as well as provide guidance towards improving it. Figure 3 gives an overview of the system components as well as the information loop.

Logue consists of two main components. First, the behavioural analysis component is responsible for perceiving, processing and classifying the user's nonverbal behaviour using various sensors. The resulting analysis is then sent to the feedback generation component where it is converted to functional icons, which are then displayed on the HMD.

Behaviour Analysis
In the scenario of public speaking, body expressivity and vocal quality are good features to measure the quality of the nonverbal behaviour of a speaker. Since the main goal of the system is to improve the user's awareness of her or his own nonverbal behaviour, we explicitly avoided any presentation-related features (e.g. total time, time-per-slide). Furthermore, using small-scale in-lab pretests, we narrowed down the feature set to features that are technically feasible to be recognized robustly in realtime. For example, while vocal clearness and loudness are good candidates, pretests showed that the analysis is very susceptible to environmental noise, the position of the microphone and non-vocal sounds (throat clearing, coughing). We ended up with three features (feedback classes) that provide the speaker with feedback on speech rate, body energy and openness.

To measure these three features, we use the Microsoft Kinect depth camera and the SHURE WH20 close-talk microphone paired to a TASCAM US 322 audio interface (Figure 4). The sensors' signals are processed by the SSI framework [37]. The SSI framework is an open-source project that supports the synchronized recording, processing and classification of sensor data in realtime. Its plugin-based architecture enabled us to easily expand SSI with new processing and classification plugins and build the signal processing pipeline required for *Logue*.

Figure 4. System setup: user wearing the HMD and microphone (far plane), and a Microsoft Kinect oriented towards him (near plane).

To compute the speech rate in realtime, we rely on work done by Jong and Wempe [10] to split voiced audio segments into syllables. The number of syllables is then divided by the length of the utterance to yield the speech rate.

Body energy is measured from the tracked positions of the user's hands in accordance with related literature [3, 7]. More precisely, we use the position of the wrist joints (as these are more robustly tracked by the Kinect than the hands) relative to the neck joint to compute the spatial displacement over the course of 5 seconds. The displacement is normalized using the arm span of the user to make the measurement user independent.

Openness is computed in a similar way. The difference is that instead of measuring the spatial displacement, we compute the Euclidean distance between the hands. However, unlike spatial extent [3], openness can also be negative if the arms cross.

Before forwarding the analysis results to the feedback component, we apply a final moving average filter to each feedback class with a window of five utterances (speech rate), 50 seconds (energy) and 30 seconds (openness). The goal here is to analyse the behaviour of the user as a whole rather than focus on fluctuations during the interaction. After this final processing step, the results of the behaviour analysis are forwarded asynchronously to the feedback generation module. In this manner, we achieve an effective update rate of one sample/utterance for speech rate and 30 samples/second for energy and openness.

Feedback Generation
The results of the behaviour analysis are converted into functional icons by the feedback generation module. Each feedback class is presented to the user using a different functional icon. A functional icon is able to express multiple intensities as well as provide information on whether the behaviour's current state is appropriate for the scenario.

For each feedback class, two thresholds (a lower and an upper threshold) determine how the behaviour state is mapped onto the intensities of the functional icons (low, medium or high). The thresholds also generate an appropriateness corridor, with behaviours within this corridor being marked appropriate and those outside inappropriate.

These thresholds were configured beforehand in a small-scale study during which, three users purposely performed well and

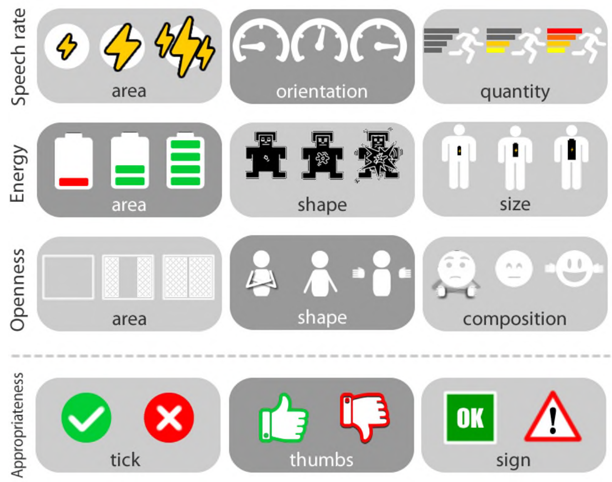

Figure 5. Initial icon set categorized by feedback classes and theme. Highlighted icon groups have been found to be most representative of their particular feedback class.

bad for each of the feedback classes. The resulting data was averaged across all participants and used to compute the two thresholds for each class. While these thresholds are meant to describe normal speaking behaviour, one could also choose other thresholds for very specific situations, e.g. to lead a presenter to perform in a highly energetic way using a loud voice.

The appropriateness of a behaviour is represented using a small symbol positioned beneath the main symbol. The main reason for separating the appropriateness from the intensity was to have both channels encoded in both shape and colour to ensure correct and fast recognition on see-through displays. Such displays are notorious for causing colour perception difficulties in uncontrolled environments [15].

Feedback Icon Selection

For *Logue* it is important that the feedback icons do not distract the users from their primary task, i.e. the social interaction itself, while still generating awareness and providing guidance. Therefore we created an initial set of 33 icons (Figure 5), designed to be understandable at a glance when viewed on a see-through HMD and to require only minimal attention for information extraction. These span our feedback classes, i.e. *speech rate*, *energy* and *openness*, and six intensity themes inspired by [4, 34] (shape, area, orientation, size, composition and quantity). The set also includes six additional icons covering the two appropriateness classes *positive* and *negative*.

Once the initial set was completed, we conducted a user study to help us select the most appropriate icons for each feedback class. For this, we split the icons into groups, each representing a single feedback class (as shown in Figure 5), and asked 25 students (4 female, 21 male; mean age of 22.4) from our university to rate on a scale from 1 (worst) to 7 (best) how representative of the feedback class each icon is.

The best rated icon groups (highlighted in Figure 5 with a dark grey background) for each feedback class have been implemented in the system. The functional icons are positioned in the upper right corner of the user's field of view to minimize distraction. The icons are shown persistently on the HMD in order to avoid unintentional jumps in gaze due to the appearance and disappearance of icons. For the visualization, we use a Vuzix STAR 1200 optical see-through HMD. It features a 1280x720 resolution spanned over two see-through displays, which offer a 23" diagonal field of view. Our system can be easily adapted to work on different HMDs, such as Google Glass.

EVALUATION

To ascertain the feasibility of the *Logue* application, we conducted two user studies using the public speaking scenario described above. For the first study, a mixed group of participants were recruited from the university to use the system in a controlled environment. They were given the task of presenting an "elevator pitch" speech to a small, targeted audience. The goal of their talks was to convince the targeted audience, which played the role of potential investors, to invest in their projects. The study aimed to quantitatively determine the impact of *Logue* on the participants' presentation performance.

For the second study, we asked senior PhD students to test *Logue* during a presentation that had to be given to peers and supervising professors at an annually organized PhD workshop. The goal of this study was to understand qualitatively the usage of *Logue* in a real setting that was not staged for the purpose of the study.

Study One: Quantitative Evaluation

The first study focuses on the collection of questionnaire data and measurement of the social signals of the participants. This allowed us to compute both subjective and objective measurements of the participants' performance.

Participants and Apparatus

We recruited a total of 15 computer science undergrad and recently graduated students (13 male and 2 female) with an average age of 26.13 (henceforth referred to as P1, $\cdots$ P15). On a 7-point Likert scale (1 = worst, 7 = very good), the participants rated themselves 3.33 for frequency of holding presentations and 4.07 on how skilful they think they are. Two employers of the institute (aged 28 and 35) were recruited to act as observers during the whole study.

The study was held in a typical conference room with the participant standing at the front of the room facing two observers at a distance of 3 m. The observers were instructed to pay attention to the nonverbal behaviour of the participants. Each participant wore the Vuzix HMD and a head-worn microphone. The Kinect was positioned in a way that did not obstruct the observers' view of the participant (Figure 6).

Task and Procedure

The participants were asked to perform two public speeches for two conditions (control and experimental) in a within subject design, i.e. one speech for each condition. In the control condition (CC) the users wore the entirety of the system setup but the feedback visualisation was deactivated. This condition provides a baseline for quantitative comparison as

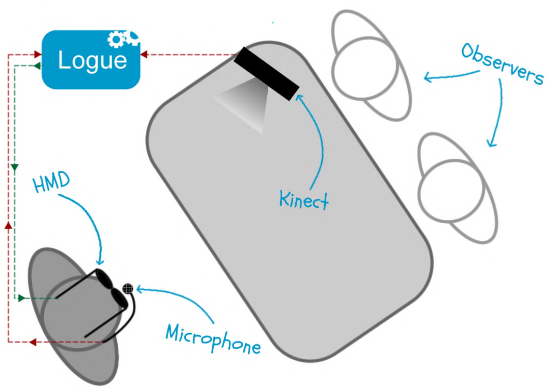

Figure 6. Evaluation setup showing the participant facing two observers while wearing the HMD and microphone. The Microsoft Kinect was positioned on the conference table between the participant and observers, and was oriented towards the participant.

no feedback was presented to the participants. In the experimental condition (EC) the participants received feedback on their nonverbal behaviour using the proposed *Logue* system. The observers were blind to how the conditions were assigned to the participants. To minimize learning effect, the two sessions were scheduled to be roughly two weeks apart and the order of the conditions was randomized. The topics of the speeches were various software projects on which the participants worked in the past half year. Each talk was expected to last approximately five minutes and no supportive materials (e.g. slides) were allowed. The participants were told to act as if the observers were potential investors, who might be interested in investing in their projects.

Prior to each session, participants filled out a questionnaire about their experience with regards to public speeches.[1] The participants were then given instructions on how to use the *Logue* system and the feedback mechanism was explained in detail. Emphasis was placed on making sure that each student had fully understood the correlation between her or his behaviour and the feedback icons. Afterwards, the participants were allowed five minutes to familiarize themselves with the system and ask any questions to the experimenters.

After each session, both the participant and the observers filled out a second questionnaire meant to elicit data regarding the participant's performance as a public speaker[1] and perceived user experience.[1] Lastly, a semi-structured interview with the user to gather general feedback about the system was conducted.

Results

We recorded video, audio, depth and social signal data from a total of 30 sessions in addition to the questionnaire data. The average length of each session was 4 minutes and 18 seconds. The audio recording from one participant had to be excluded from the analysis due to a medical condition that caused frequent throat clearings and interfered with the audio analysis. Similarly, the recording of one participant was omitted from the energy and openness analysis due to skeleton tracking problems.

[1]`http://hcm-lab.de/downloads/chi15`

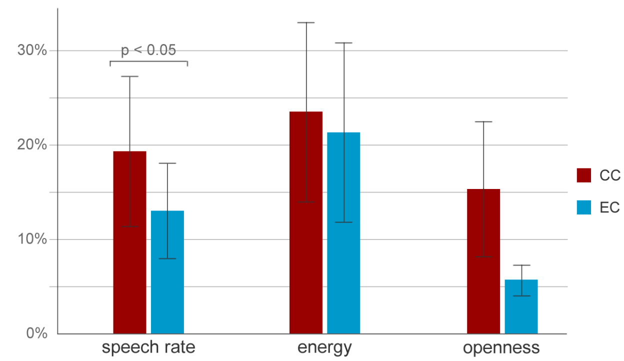

Figure 7. Percentage of inappropriate behaviour (y-axis) for each feedback class across conditions (control vs. experimental). Lower values are better.

(A) Impact on objective measurements. We processed the recorded data for each feedback class by computing how many of the participants' vocal segments ("utterances") measured speech rates outside of the thresholds. We normalized this value by the total amount of utterances of the session and averaged it over all participants to get a measure of the speech rate inappropriateness for each condition. We proceeded in an analogue way for energy and openness. However, since the energy and openness are computed continuously (unlike the speech rate), we used the normalized duration in seconds of the time spent outside of the thresholds instead of the number of utterances.

The resulting values can be seen in Figure 7. A One-Way Repeated Measure MANOVA revealed by trend a multivariate effect of social augmentation on the amount of inappropriate behaviour for speech rate, energy and openness, $F(3, 11) = 3.215$, $p = .065$. When looking at the feedback classes individually, a Wilcoxon signed rank test showed significant differences between the conditions for speech rate with $p = 0.033$, $Z = -2.134$. There was no significant effect for energy and openness though. As Figure 7 illustrates, we measured increased standard errors of the mean for all feedback classes. These were caused by the participants who never crossed the thresholds and thus yielded zero amount of inappropriate behaviour.

When taking a closer look at the data we can observe how the participants reacted to the feedback. Figure 8 shows P13's openness over time and how it was affected by the behavioural feedback (symbolized by the vertical lines). As can be seen in the figure, P13 immediately reacted on the systems feedback by performing more open gestures.

(B) Impact on observer's perception. The observers rated the participants as significantly more open ($F(1, 14) = 3.333$, $p = 0.045$) when the social augmentation mode was activated than when it was not. This suggests that the social augmentation mode had an observable impact on the participants' behaviour. However, we did not find any significant differences for the other dimensions of the observers' questionnaire.

(C) Impact on self perception. T-tests for one sample revealed that the participants' subjective ratings of the social augmentation mode were significantly above the neutral value of 4.0

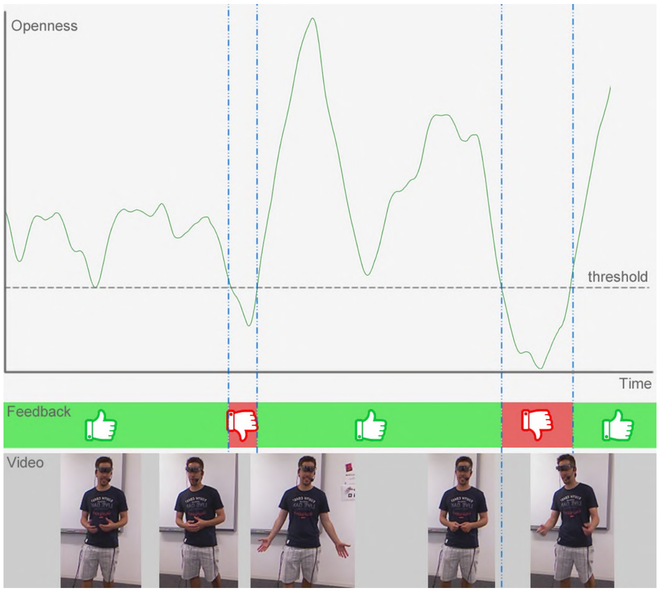

Figure 8. Example of participant's openness over the course of a session in the experimental condition.

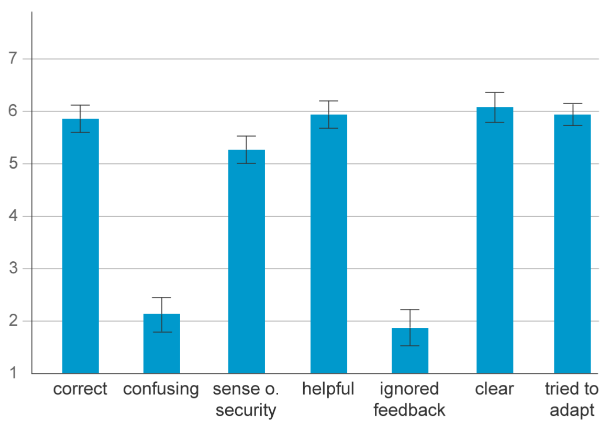

Figure 9. Results of the user experience questionnaire showing means on a 7-point Likert scale (1 = worst, 7 = very good). Two items (*confusing* and *ignored feedback*) are reverse-scored.

(Figure 9). The participants thought the feedback was correct ($t(13) = 6.77$, $p < .0005$), helpful ($t(14) = 7.25$, $p < .0005$) and not confusing ($t(14) = -5.332$, $p < .0005$). The feedback also gave them a sense of security ($t(14) = 4.75$, $p < .0005$) and they said that they did not try to ignore the feedback ($t(14) = -6.094$, $p < .0005$). A comparison between the conditions yielded no evidence that the participants felt distracted by the social augmentation mode or that the social augmentation mode increased the difficulty of the main task.

Study Two: Qualitative evaluation in a real setting

Following our first study, where we made use of an enacted scenario, we conducted a second study in order to test *Logue* in a real presentation setting. To this end, we recruited speakers from an annual doctoral workshop who volunteered to make use of *Logue* during their presentation. The workshop gives computer science PhD students the opportunity to present the current state of their work to their peers and supervisors. The experimental setting did not only expose the participants to more realistic conditions, but the quality of the presentation also had an impact on the participants' standing within the laboratory. Apart from the fact that some people made use of *Logue* during their presentation, the workshop was held in a similar manner as the years before.

Participants and Apparatus

For the second study, three senior computer science PhD students (P16, P17 and P18) were recruited. Similar to the other workshop participants, they had to give a prepared presentation on their PhD topics. The talks took place in a seminar room, with the participant standing at the front of the room and facing the audience, which consisted of 13 peers and two supervising professors. All speakers made use of slides to accompany their presentation. The presentations lasted about 30 minutes and typically included ten minutes of discussion.

Task and Procedure

After the talk and the scientific discussion, an open discussion on the style of presentation followed. The audience was asked questions regarding the perceived quality of the talk as well as whether they felt the proposed system had influenced the quality of the presentation in a positive or negative manner. Later, we conducted a semi-structured interview with each participant to elicit her or his impression of the system.

Results

We highlight three outcomes of the semi-structured interviews regarding *Logue*'s impact on behaviour, its level of distraction and its usefulness.

Did you adapt your behaviour? Speaker 1 (P16) stated "One time during the presentation I felt I was talking too fast and then I remembered I had this thing on my nose so I looked at the feedback and this was actually the case. I then tried to talk slower." We then asked the audience what they thought about the speech rate. One observer said "[the speech rate] was not actually disturbing, but it would have helped if he would have talked slower," indicating that there were indeed issues with the speaker's speech rate. P17, who told us prior to the study that he is aware he talks really fast, explained: "I was surprised that the speech rate did not become red sooner [...] once I saw the feedback that I was talking too fast, I tried too adapt." Hence, *Logue* had an effect on their presentation and they tried to adapt their behaviour as suggested by the system.

Was the system distracting? All participants complained about the bulkiness of the HMD. Despite this, P16 added "it is a very interesting concept, most of the time I did not perceive the system, only when I consciously looked at the feedback. It would be interesting to try it out for a longer period of time, for example to use it when teaching a class ... but with the Google Glass." P17 also said "the system was unobtrusive [...] I consciously looked at the feedback from time to time." We take this as a positive sign that *Logue* does not pose an unacceptable distraction for the speakers.

Would you use it? When asked whether they would use the system, all were positive. P16 said: "I would use it during real presentations or while teaching a class. Or to train for a presentation;" P17 was more reserved and said "If you used it regularly you would get a feeling for what's good or bad so that you might not need it any more after a while." These statements are encouraging, suggesting that *Logue* provides sufficient value to warrant its regular use, for example during lectures.

Discussion

Impact on behaviour

The results from the first study suggest that the system has an impact on the behaviour of the participants. For all three feedback classes, we were able to measure improvements over the control condition in terms of amount of appropriate behaviour as measured by the system, with significant effects for the speech rate.

To get a better understanding of how the system impacted the behaviour of a speaker, we refer to Figure 8. It shows how P13 adapted his openness after the feedback changed, but also how it slowly degraded over time until another correction was needed. This effect further strengthens the potential of the proposed approach, as it acts similar to a reinforcement technique, repeatedly reminding the users to adapt their behaviours. This effect was found among multiple participants and is backed up by statements from the semi-structured interviews from both studies. For instance, P1 stated "I think the system helped me a lot as it reminded me to talk slower. This is a problem many people pointed out to me after I hold a presentation and it felt good to get the feedback during the talk."

In some cases we found evidence of a longer term effect on the participants. During his second session (CC), P10 used very broad gestures when describing his project's composition. When confronted, he admitted to remembering the feedback from the first session (EC) despite them being two weeks apart. While this effect may have had a negative impact on the actual goal of the first user study, it is nevertheless encouraging for the scope of the project and we plan to investigate this as part of our future work.

Need for personalization

It is interesting to note that some participants did not cross the thresholds at all. Furthermore, participants seemed to react differently to the feedback. Some adapted the behaviour instantly, others gradually and a few participants ignored it. This can be explained by a heterogeneous distribution of presentation skill among the participants. If we classify the participants into experts and novices based on their answers from the demographics questionnaire, we notice some interesting trends. Novice users improved their speech rate more than experts (mean decrease in amount of inappropriate behaviour $\Delta = 9.3\%$ versus $\Delta = 0.9\%$). This shows that experts are already speaking at an optimal speech rate. Remarkably, when looking at the openness dimension, the effect is reversed. The experts managed to improve more ($\Delta = 31.5\%$) than the novices ($\Delta = 1.2\%$). A possible explanation for this is that the participants who rated themselves bad at public speaking (i.e. the novices) appeared to be more introverted and thus had difficulties performing open gestures even when given feedback to do so. For example, during the post-hoc interview, P1 said "I did not know how to correct my behaviour. I tried to perform more gestures, but I found it very difficult to incorporate these into my performance." A similar statement came from P15: "I knew I had to move more, but I was afraid that simply moving my hands would look weird."

We argue that these differences denote the importance for a customizable system. More precisely, there is need for not only scenario-specific customization but also user-specific individualization. This is backed up by another observation during the post-interaction interviews: despite all participants receiving equally detailed feedback, some participants stated that they were overwhelmed by the feedback whereas others asked if we can increase the level of detail on the behaviour analysis. We received similar feedback from the second study, with P16 stating: "It would be good to know how far above the threshold I am." P17 also noticed the need for customization when saying that in its current state, *Logue* "is good for newbies" but "probably not accurate enough for professionals."

Participants also stated that they can envision themselves using the system as preparation for stressful social situations such as public speeches. In such training environments, an increased level of detail for the peripheral feedback would be feasible, as the primary task is no longer the actual interaction, but the training itself. We intend to investigate this aspect in future evaluations.

Not a distraction

When asked whether the system distracted them from the act of giving a talk, the participants assured us that the impact was minimal. P15 stated "It felt distracting at first, but then I noticed that I can look through the displayed feedback. Once this was clear, I only glanced at the feedback from time to time to see whether anything has changed." Other participants, including the senior PhD students from the second study, made similar statements, saying that they were checking on the behaviour during speech or thinking pauses to see if something had changed. These results support our initial design choices, in particular the use of persistent visual feedback. We argue that in this manner, the participant is able to decide when to access the information, thus minimizing the impact on the primary task. We aim at a more in depth analysis of this aspect in future work. We are especially keen on comparing different feedback delivery methods and modalities.

Sense of security

The semi-structured interviews also revealed an increased sense of security for the participants in the experimental condition. This was even the case for participants who behaved appropriately the whole time. P6 expressed: "I would look at the icons from time to time and seeing them green made me feel better about my performance." The participants who did the control condition after the experimental condition even stated that they "missed" the system as they were unsure of the appropriateness of their behaviour. We received similar statements from the three senior PhD students in the second study, who can be regarded as more experienced speakers. Both P17 and P18 mentioned an increased sense of security, with P18 saying "It was a good feeling seeing everything green ... it's like applause, or as if someone looks at you and nods. However, the green lasts longer than a nod [laughs]." P17 was more cautious: "it could help with feeling more secure [...] if you are untrained and get cold feet."

Social acceptability

During the second study we also asked the audience after every presentation for their thoughts and feelings on the system. Overall, they mentioned that the system appeared odd to them in the beginning because of the bulkiness of the HMD. However, they soon got used to it and it did not distract them from the actual presentation. It is very likely that a lighter HMD would increase the acceptability in the eyes of the audience. No privacy-related concerns were raised by the audience or the speakers themselves. However, this is most likely due to the study setup and the technical affinity of the participants. Nevertheless, we argue that as long as the system only analyses the user's own behaviour, *Logue* does not violate the privacy of surrounding persons.

Limitations

A major limitation of this initial implementation of augmented social interaction was the HMD, as almost all participants complained about its weight and size. However, these issues will be mitigated when transitioning to a lighter HMD, such as Google Glass.

Although the system had a positive effect on the participants' behaviour as measured by the system, no major effects were noticeable on the perception of the observers. One possible explanation for this is that accurately assessing another person's behaviour is very difficult. Furthermore, the quality of a presentation is also a very subjective measure making the search for an objective definition of a perfect presentation futile. Most of us have witnessed good talks, but which were nothing alike. Considering this, it is possible that for our observers, different feedback classes and thresholds might have had a larger effect on their perception of the talks. The fact that we only had two observers (the objective sensor measurements were the study's main focus) is also likely to have contributed. In future studies we would like to employ a larger number of observers in the hope of getting a better assessment of the audience's perception.

CONCLUSION AND FUTURE WORK

The aim of augmented social interaction is to inform the user of shortcomings in the nonverbal behaviour and thus, enable her or him to adapt the behaviour in an attempt to improve the outcome of the interaction. To get a better understanding of how this would translate to an actual system, we implemented *Logue*, a system that augments public speakers using a head-mounted display and various sensors. *Logue* provides the user with behavioural feedback on two levels. First, it informs the user of the current state of her or his speech rate (How fast am I talking?), body energy (How much do I gesticulate?) and openness (How open is my posture?). Secondly, the system indicates the quality of each of these three behavioural cues in relation to the public speaking context. For example, a high speech rate would be marked as inappropriate as it may impact the listeners' ability to follow the talk. We evaluated *Logue* not only in a staged, but also in a real presentation setting to see how it impacts the behaviour of the user. Both studies yielded promising results, with users being positive about the effect the system has on one's performance and mental state while not distracting from the main task. In the first study we also collected objective data from the behaviour analysis component. Here we found evidence for improvements across all three dimensions, i.e. speech rate, body energy and openness, when using *Logue*.

The main contribution of the paper lies in the design and implementation of a new form of realtime nonverbal behaviour feedback during social interactions using wearable devices and sensors. To achieve this, multiple software components have been designed and integrated into one coherent system. The system is open source and available for download.[2]

Overall, we find it highly motivating to see how well this initial prototype was received. Its mix of peripheral feedback and social signal processing enabled users to adapt their nonverbal behaviour during the actual interaction without degrading the primary task. Work is still required on the scenario level, especially when it comes to choosing feedback classes and setting thresholds. Here, we would like to conduct more in-depth studies to accurately determine which configuration of feedback classes and thresholds are representative for good and which for bad public speaking behaviour.

Acknowledgements. This work was partially funded by the EC within FP7-ICT-2011-7 (288578 - TARDIS)

REFERENCES

1. Batrinca, L., Stratou, G., Shapiro, A., Morency, L.-P., and Scherer, S. Cicero - towards a multimodal virtual audience platform for public speaking training. In *Proc. IVA*, vol. 8108. Springer, 2013, 116–128.

2. Baur, T., Damian, I., Gebhard, P., Porayska-Pomsta, K., and André, E. A job interview simulation: Social cue-based interaction with a virtual character. In *Proc. SocialCom*. IEEE, 2013, 220–227.

3. Baur, T., Damian, I., Lingenfelser, F., Wagner, J., and André, E. Nova: Automated analysis of nonverbal signals in social interactions. In *Human Behavior Understanding*, vol. 8212 of *LNCS*. Springer, 2013.

4. Bertin, J. *Semiology of graphics: diagrams, networks, maps*. University of Wisconsin press, 1983.

5. Birdwhistell, R. L. *Kinesics and context: Essays on body motion communication*. University of Pennsylvania press, 2011.

6. Blanch, D. C., Hall, J. A., Roter, D. L., and Frankel, R. M. Is it good to express uncertainty to a patient? correlates and consequences for medical students in a standardized patient visit. In *Proc. EACH*, vol. 76 (2009), 300 – 306.

[2] `http://goo.gl/QXGyRG`

7. Caridakis, G., Raouzaiou, A., Karpouzis, K., and Kollias, S. Synthesizing gesture expressivity based on real sequences. *Multimodal Corpora* (2006).
8. Carl, H. Nonverbal communication during the employment interview. *ABCA Bulletin 44*, 4 (1980).
9. Cohen, B. M., and Etheredge, J. M. Recruiting's main ingredient. *Journal of College Placement* (1975).
10. de Jong, N. H., and Wempe, T. Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods 41*, 2 (2009), 385–390.
11. DeVaul, R. W., Pentland, A. S., and Corey, V. R. The memory glasses: Subliminal vs. overt memory support with imperfect information. In *Proc. ISWC* (2003).
12. Drake, L. R., et al. How do employers value the interview?. *Journal of College Placement 32*, 3 (1972).
13. Engelbart, D. C. Augmenting human intellect: A conceptual framework. Air Force Office of Scientific Research, AFOSR-3233, 1962.
14. Fukumoto, K., Terada, T., and Tsukamoto, M. A smile/laughter recognition mechanism for smile-based life logging. In *Proc. AH*, ACM (2013), 213–220.
15. Gabbard, J. L., Swan, J. E., Zedlitz, J., and Winchester, W. W. More than meets the eye: An engineering study to empirically examine the blending of real and virtual color spaces. In *Proc. VR*, IEEE (2010), 79–86.
16. Hill J, R. W., Gratch, J., Marsella, S., Rickel, J., Swartout, W. R., and Traum, D. R. Virtual humans in the mission rehearsal exercise system. *KI 17*, 4 (2003), 5.
17. Hogan, K. *Can't Get Through: Eight Barriers to Communication*. Pelican Publishing, 2003.
18. Hollandsworth, J. G., Kazelskis, R., Stevens, J., and Dressel, M. E. Relative contributions of verbal, articulative, and nonverbal communication to employment decisions in the job interview setting. *Personnel Psychology 32*, 2 (1979), 359–367.
19. Hoque, M. E., Courgeon, M., Martin, J., Mutlu, B., and Picard, R. W. Mach: My automated conversation coach. In *Proc. UbiComp* (2013).
20. Hwang, I., Yoo, C., Hwang, C., Yim, D., Lee, Y., Min, C., Kim, J., and Song, J. Talkbetter: Family-driven mobile intervention care for children with language delay. In *Proc. CSCW*, ACM (2014), 1283–1296.
21. Kim, T., Chang, A., Holland, L., and Pentland, A. S. Meeting mediator: Enhancing group collaborationusing sociometric feedback. In *Proc. CSCW*, ACM (2008).
22. Lee, Y., Min, C., Hwang, C., Lee, J., Hwang, I., Ju, Y., Yoo, C., Moon, M., Lee, U., and Song, J. Sociophone: Everyday face-to-face interaction monitoring platform using multi-phone sensor fusion. In *Proc. MobiSys*, ACM (2013), 375–388.
23. Li, X., and Rekimoto, J. Smartvoice: a presentation support system for overcoming the language barrier. In *Proc. CHI*, ACM (2014), 1563–1570.
24. Lu, H., Frauendorfer, D., Rabbi, M., Mast, M. S., Chittaranjan, G. T., Campbell, A. T., Gatica-Perez, D., and Choudhury, T. Stresssense: Detecting stress in unconstrained acoustic environments using smartphones. In *Proc. UbiComp*, ACM (2012), 351–360.
25. McAtamney, G., and Parker, C. An examination of the effects of a wearable display on informal face-to-face communication. In *Proc. CHI*, ACM (2006), 45–54.
26. Mehrabian, A. *Silent messages: Implicit Communication of Emotions and Attitudes*. Wadsworth Publishing Co Inc, Belmont, 1981.
27. Ofek, E., Iqbal, S. T., and Strauss, K. Reducing disruption from subtle information delivery during a conversation: Mode and bandwidth investigation. In *Proc. CHI*, ACM (2013), 3111–3120.
28. Pan, X., Gillies, M., Barker, Clark, D. M. C. M., and Slater, M. Socially anxious and confident men interact with a forward virtual woman: An experiment study. *PLoS ONE 7*, 4 (2012), e32931.
29. Pease, A. *Body Language*. Sheldon Press, London, 1988.
30. Posthuma, R. A., Morgeson, F. P., and Campion, M. A. Beyond employment interview validity: A comprehensive narrative review of recent research and trends over time. *Personnel Psychology 55*, 1 (2002).
31. Saket, B., Yang, S., Tan, H., Yatani, K., and Edge, D. Talkzones: Section-based time support for presentations. In *Proc. MobileHCI*, ACM (2014), 263–272.
32. Scherl, C. R., and Haley, J. Computer monitor supervision: A clinical note. *The American Journal of Family Therapy 28*, 3 (2000), 275–282.
33. Sieverding, M. be cool!: Emotional costs of hiding feelings in a job interview. *International Journal of Selection and Assessment 17*, 4 (2009), 391–401.
34. Szczerba, J., Hersberger, R., and Riegelman, A. Design and evaluation of a differential speedometer. In *Proc. HFES Meeting*, vol. 56 (2012), 1629–1633.
35. Tan, C. S. S., Schöning, J., Luyten, K., and Coninx, K. Investigating the effects of using biofeedback as visual stress indicator during video-mediated collaboration. In *Proc. CHI*, ACM (2014), 71–80.
36. van der Linden, J., Johnson, R., Bird, J., Rogers, Y., and Schoonderwaldt, E. Buzzing to play: Lessons learned from an in the wild study of real-time vibrotactile feedback. In *Proc. CHI*, ACM (2011), 533–542.
37. Wagner, J., Lingenfelser, F., Baur, T., Damian, I., Kistler, F., and André, E. The social signal interpretation (ssi) framework - multimodal signal processing and recognition in real-time. In *Proc. ACM MM* (2013).
38. Xia, C., and Maes, P. The design of artifacts for augmenting intellect. In *Proc. AH*, ACM (2013).