Dites-Moi: Wearable Feedback on Conversational Behavior

Skanda Muralidhar
Idiap Research Institute
EPFL, Switzerland
smuralidhar@idiap.ch

Jean M. R Costa
Cornell University, U.S.A
jmd487@cornell.edu

Laurent Son Nguyen
Idiap Research Institute
EPFL, Switzerland
lnguyen@idiap.ch

Daniel Gatica-Perez
Idiap Research Institute
EPFL, Switzerland
gatica@idiap.ch

ABSTRACT

Interpersonal communication skills are critical in certain industry sectors like sales and marketing. Recent advances in wearable technology are enabling the design of real-time behavioral feedback tools for apprentices in aforementioned industries. This paper describes the design and implementation of a conversational behavior awareness tool based on Google Glass. The goal of the system is to provide real-time feedback to young sales apprentices about the amount of time they talk in an interaction with a client. We evaluated our system with a pilot study involving 15 apprentices (ages 16-20). Overall, participants found the system fun, little distracting and useful. Furthermore, manual coding of the recorded videos, showed that wearable sensing and real-time feedback did not negatively influence the dyadic social interaction.

Author Keywords

Nonverbal Behaviour, Real-Time Feedback, Wearable Devices, Ubiquitous Computing, Google Glass

INTRODUCTION

Interpersonal communication is the sine qua non of social interactions as it represents the means through which we initiate, negotiate, and maintain human relationships [16]. Hence, interpersonal skills are paramount in the context of workplaces and are critical in certain sectors like hospitality, sales and marketing. This has major implications for the quality of human resources and specifically for training and development of interpersonal communication skills. In this work, we present a wearable application developed for Google Glass (GG) that can provide behavioral awareness for young apprentices of vocational education and training (VET) school.

Literature in psychology has demonstrated that nonverbal behavior is a major channel of interpersonal communications [3, 13] . The widespread availability of inexpensive sensors combined with improved perceptual techniques have enabled the possibility to automatically analyze social interactions [7, 21]. In the context of workplaces, recent studies have established the feasibility of automatically inferring interview ratings [18], negotiation outcomes [4] and other related constructs (e.g. engagement, friendliness, or excitement) [11, 22, 17] up to a certain level. Ensuring improved behavioral skills of the apprentices is an important task in VET schools where sales and customer service are taught.

Existing psychology literature indicates that social interaction skills can be improved by practicing both verbal and nonverbal communication including how much, how fast, and how loud to talk, and how to regulate turn taking [10]. Advances in ubiquitous and wearable computing are enabling new possibilities to deliver real-time feedback [9, 19] and uses in the classroom, like physics experiments [26].

Providing real-time feedback during conversations has been investigated in the past. In the context of group interactions, feedback to participants was provided by projecting their speaking time on a large common surface like a wall [6] or on a customized table which acted as both sensing and display platform [1]. A mobile phone-based solution for sensing and displaying a person's nonverbal cues (speaking time, prosody and body movements) was developed in [12], yielding a reduction of behavioral differences between dominant and non-dominant participants. In [24], feedback systems that combine visual and acoustic cues, e.g. automatically estimating speaking time and visual attention using headbands tracked by infrared camera were developed.

In the context of public speaking, GG has been used as a head-mounted display system to provide real-time feedback on a presenter's posture openness, body energy, and speech rate sensed using data provided by Kinect and an external microphone [5]. In [25], GG has been used to display information and as an audio sensor to provide automatic real-time feedback on a speaker's speaking rate and energy. The data was processed on an external server.

In the context of dyadic interaction, [15] investigated the effect of a head mounted device on social interaction. The authors reported a degradation of social interaction and eye contact. However, display on the screen was a series of slides showing emails, text messages etc with each slide being visible for 40 seconds. We see an opportunity in the design of tools using GG that provide real-time feedback during face-to-face interaction

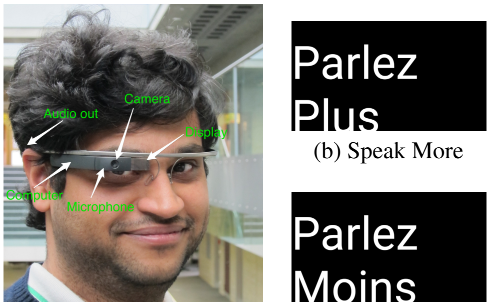
Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

MUM '16, December 12 - 15, 2016, Rovaniemi, Finland

© 2016 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ISBN 978-1-4503-4860-7/16/12...$15.00

DOI: http://dx.doi.org/10.1145/3012709.3012733



(a) Glass Sensors (b) Speak More (c) Speak Less

Figure 1: Overview of GG sensors and visual feedback on GG Display

to increase self-awareness of conversational behavior, while not impairing the quality of interaction.

The objective of this work was to design and implement an automatic, real-time, conversational behavior awareness system for young sales apprentices to make them aware of their nonverbal behavior while interacting with a customer. Towards this objective, we designed and developed a pilot GG app which provides real-time behavioral feedback. We evaluate the design and usefulness of the app by conducting a pilot study with 15 sales apprentices from a VET school. We believe that some of the results obtained in this study can also be applied to other devices capable of displaying behavioral feedback, such as smart watches, tablets, etc.

The contributions of this work are as follows. First, we designed and developed a GG app for providing real-time feedback on speaking time. Second, we collected a dataset of 15 sales interactions with apprentices recorded in the lab, where GG was used as a self-awareness tool. Third, we demonstrated the usefulness of this tool and analyzed the effect of real-time feedback on conversation. Our work constitutes a first step towards designing a real-time behavioral training tool in a dyadic setting that is appealing to young apprentices in jobs involving positive interpersonal skills.

APPROACH

Providing feedback during a dyadic interaction without negative impact is a challenge. The main challenge is to provide speakers with key behavioral insights without distracting them from their interaction. The human brain is not adept at multitasking [20], hence any significant distraction might lead to behavioral artifacts like stuttering, awkward pauses or smiles. Additionally, by continuously staring at the feedback screen, the speaker might lose eye contact with the protagonist, causing the quality of interaction to degrade.

Considering these constraints, GG was utilized as a display to provide feedback. The GG screen is a small high resolution display located at the periphery of user's field of vision and engineered to have minimal cognitive load. For aural feedback, the bone conduction transducer was utilized. Using this technology, audio is sent directly to the inner ear through bones of the skull, rendering it audible only to the user.

The behavioral tool consists of two main components; sensing and feedback. The sensing component is responsible for perceiving and processing the user's nonverbal behavior. The built-in microphone of GG was exploited towards this end (Fig. 1a). The feedback generation component uses the resulting analysis and presents the appropriate messages either visually or aurally. The following section details both the components of our behavioral awareness tool and an user case study conducted to verify the effectiveness of the said tool.

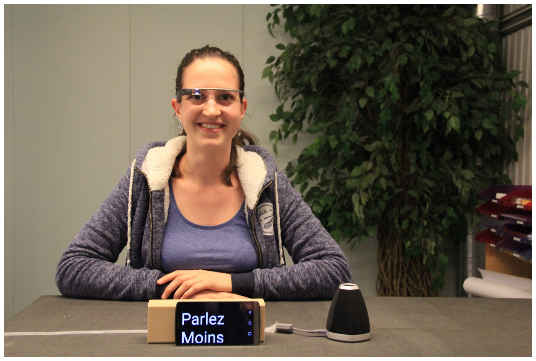


Figure 2: View of the study setting: the participant works behind the desk; the interaction partner is not visible in the figure.

App Implementation

A prototype of the behavioral awareness tool was designed and implemented on two GG devices using the Android platform. The devices were running Android OS and implementation of the application was done in Java.

A significant effort was devoted to evaluate what nonverbal features should be shared with the user. Literature in psychology and social computing indicate several nonverbal cues to be important during a dyadic interactions in the context of workplaces [13, 18]. However, due to constraints of low computational power of GG, nonverbal cues with moderate computational requirements were considered. In an initial design phase, speaking time and a proxy of head orientation were considered and implemented. Hence, feasibility of estimated gaze and speaking time was investigated.

A small pre-trial was conducted with three lab colleagues to evaluate the experience of this initial design. The results of the pilot study showed that the use of the two nonverbal cues (speaking time and gaze) significantly affected the duration of GG battery and lead to heating of the device to uncomfortable levels. Another reason that necessitated a simple interface was the sample population in the evaluation use case. Participants in the user study reported themselves to be inexperienced with wearable devices (mean$=1.8$, median$=2$) and reluctant to use new technology (mean$=3.1$, median$=3$) (scale $1-7$). Given these factors, the final design was focused on speaking time, which is intuitive to users engaged in conversation, and is backed up by literature in psychology as a cue related to extraversion and dominance among other constructs [13].

To compute speaking status, speech captured by the built-in microphone of GG was utilized. The speech non-speech segmentation was performed using a two-step approach. First, the subject's voice was segmented from the other protagonist using audio energy as a discriminative feature: the microphone is significantly closer to the subject than the other interlocutor, therefore the subject's voice is assumed to be louder. Second, we used the method proposed by [2], which was shown to be robust in noisy environments [14]. The method is independent to energy and uses a two-layer binary HMM: the low-level latent variable is voiced/non-voiced and the high-level one is speech/non-speech. The processing was done on GG itself.

The sensing component provides analysis for the feedback component. Our prototype currently furnishes feedback based on a window of 20 seconds. If no voice activity is detected for 20 seconds, GG prompts the user to speak. If continuous voice activity is detected for more than 20 seconds, the tool prompts the user to stop talking. Although we acknowledge that this 20-second threshold can be somewhat arbitrary, this duration was chosen based on detailed discussions with colleagues in psychology and its use in existing literature [25]. Additionally, the objective of this study was not to investigate the best speaking duration; rather, we focused our analysis on the effect of feedback on the quality of the interaction.

Feedback on speaking time was provided as one of two possible modalities: visual and aural. For both modalities, the feedback was not noticeable by the other interlocutor. Visual feedback was provided using text, which was presented to the user sparsely as suggested by authors in [25]. The text, screenshot in Figures 1b & 1c, prompted the participants to '*speak less*' or to '*speak more*'.

Aural feedback, a modality that has been used less often in the social sensing literature, was provided in the form of prerecorded speech ('*speak less*','*speak more*'). The bone conduction transducer was utilized to provide this feedback.

Scenario

To evaluate the usefulness of the system, we conducted a user study with 15 participants. Subjects were volunteers from a local VET school, who participated as an opportunity to improve their communication and sales skills. Of the 15 students, 9 were male. Average age was 17.7 years old. The subjects reported little professional sales experience (mean= 1.75, median= 1 on a $1-7$ Likert scale). They were randomly split with one half provided with visual feedback while the other group was presented with audio feedback.

The interaction consists of a typical sales scenario in a mobile phone shop (average duration = 2.5 minutes). In this scenario, the participant played the salesperson role. Each student had to interact with a customer with the goal to satisfy the client, and try to sell them the best (= most expensive) data package along with the phone (iPhone). During the interaction, GG would provide automatic feedback on behavioral cues. It was in their discretion to follow the suggestion or not. The role of the client was played by a researcher who was a native French speaker with directions to elicit two behaviors from the participants: talk for a relatively long time, and remain silent. A snapshot of the scenario is presented in Figure 2. All interactions (for both partners) and GG feedback are video recorded with Kinect devices (Fig. 2).

EVALUATION

To gain insight on number of times feedback was given and the type of feedback provided, the videos were manually coded by the authors. Due to the design of the experiment, each participant received feedback at least once. It was further observed that three participants received feedback twice. Also, three participants received feedback to '*speak more*, while other received feedback '*speak less*'.

Table 1: List of questions in the self reported pre- and post-questionnaires rated on a Likert scale

Pre-Questions (Self Reported)	Post-Questions (Self Reported)	
	Visual	Audio
Sales Experience	Usefulness	Usefulness
GG Experience	Distracting	Distracting
Interest in GG	Overall Impression	Overall Impression
Interest in Technology		

To understand the user's perspective on using GG, the app, and to identify issues with current prototype implementation, evaluation was carried out by analysis of participant self-reported questionnaires and external annotator impressions.

Questionnaire Data

The participants were asked to fill two questionnaires, one before and the other after the interaction. In both questionnaires, subjects had to rate various questions on a Likert scale (where 1 ='very poor'; 7 ='very good'). The list of questions asked in both the questionnaires are presented in Table 1. Additionally, the pre-questionnaire consisted of demographic details and a personality test. The personality test, in French, was administered using a Ten Item Personality Inventory (TIPI) [8]. Analysis of personality is planned as part of future work. The post-questionnaire required subjects to answer only for the modality they were presented during the data collection.

Figure 3a shows the self rated impressions of participants for both modalities of feedback. It can be observed that participants find the feedback using audio modality to be useful (median = 4) but find it to be a little distracting (median = 3). On the other hand, participants who were given visual feedback found this modality to be less distracting than audio (median = 2) and more useful (median = 5). A possible explanation for this difference is that the audio feedback could have interfered with the speech of the protagonist. The participants reported a positive overall impression for both modalities (median = 5).

Broadly, the participants indicated a positive overall impression towards an wearable behavioral feedback tool (Figure 3b). They also indicated that the wearable device and app were found to be Natural (median = 3.5), Cool (median = 5.5), Comfortable (median = 5) and Fun (median = 5) during the dyadic interaction. Thus, the results indicate that subjects found the real-time behavioral feedback to be useful, natural and comfortable. These results are in line with those reported in literature [5, 25], and are novel from the perspective of the specific use of GG by a very young population. At the end of data collection, in an informal discussion with the participants, they all expressed the usefulness of the app. In particular, participants favored visual feedback over audio feedback.

External Observer Annotations

To assess the impact of glass on dyadic interaction, the sales video was annotated by two groups of native French speakers. Group-A and Group-B consisted of two and three raters respectively. Group-A was informed, at length, about GG and the feedback provided by it, while Group-B was not. For both groups, the part of the screen which displayed feedback was

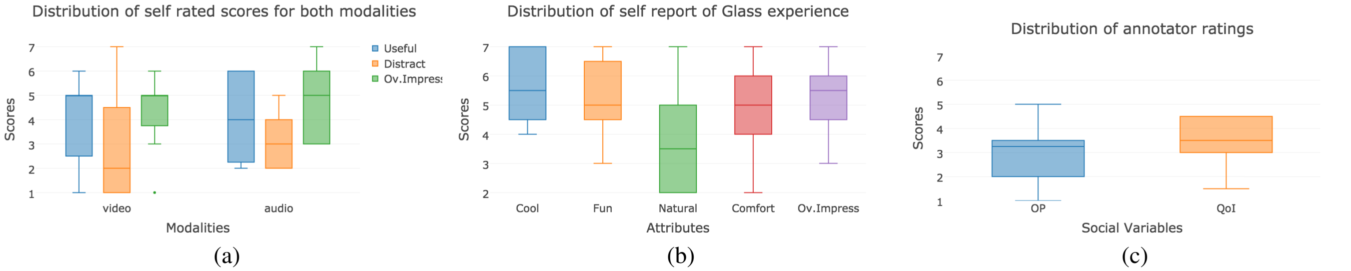


Figure 3: Distribution of participants' self ratings for (a) feedback modalities (higher is better for *useful* and *Ov. Impression*, lower is better for *Distract*); (b) overall experience (higher is better); (c) *overall performance* and *quality of interaction* ratings by annotators (higher is better)

blocked. Group-A was asked to watch the video and answer: *Do you believe the salesperson was given feedback, based on the behavior of the person throughout the video?* in the form of *Yes*, *No* or *Maybe*. Annotators in Group-B were asked to rate the video on a five-point Likert scale (1 ='very poor' to 5 ='very good'). Specifically they were asked *Consider yourself to be the client in this interaction and rate* (a) *Overall performance (OP) of the participant* (b) *Quality of interaction (QoI) with the participant*.

Agreement between the raters in Group-B was calculated using Intraclass Correlation Coefficient (*ICC*), as a measure commonly used in psychology and social computing [23]. $ICC(2,k)$ was used as all raters gave scores for each video. The obtained ICC values were above 0.70 for both the social variables [OP: $ICC(2,k) = 0.90$, $p < .001$; QoI: $ICC(2,k) = 0.70$, $p < .001$]. This indicates that the agreement between raters was high for both social variables. Final scores for both social variables were obtained by taking the mean of all scores.

The distribution of annotation data for both variables is presented in Figure 3c. Median rating of OP is 3.25 (max= 5; min= 1), while the median rating of QoI is 3 (max= 4.5; min= 1). Due to the limitations in dataset size, no firm statistical conclusions can be drawn for social variables. Also, talking more does not imply a better conversation. Another limitation of this work is that the QoI and OP was not validated by domain experts (speaking coach or sales coach).

Figure 4 indicates that for majority of the videos, the annotators of Group-A were unable to correctly infer if feedback had been provided (adding the "no" and "maybe" columns in Figure 4). These results suggest that in several cases the reaction of GG users to the feedback is either subtle or does not deviate from what an external observer would consider as usual conversational behavior.

To investigate this issue in more detail, the behavior of participants during the interaction was manually coded by the authors to understand how subjects react to real-time feedback. The manual coding of behavior signal that some subjects smiled or giggle when feedback was provided, possibly due to both the actual experience of receiving feedback combined with a novelty effect. Reactions to both types of feedback and time to heed to suggestion is presented in Table 2. This in conjunction with annotations by Group-A on inference of feedback (Figure 4) indicate that in the majority of the cases reaction to feedback was natural.

Table 2: Behavioral reactions to feedback. Time to heed is the time to taken to accept the feedback i.e stop talking if feedback says stop talking.

Feedback Type	Reaction	Time to heed
Speak Less	Smiling, Laughing, Squinting	1-4 seconds
Speak More	Smiling	2-4 seconds

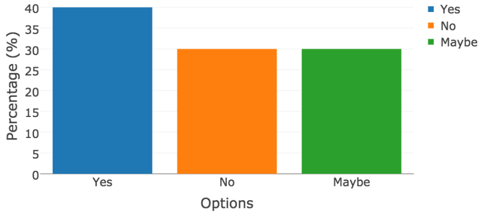


Figure 4: Distribution of answers for prediction by Group-A. All GG users received feedback.

CONCLUSION

This paper presented the design and evaluation of a real-time wearable prototype for self-awareness of conversational behavior, aim to support young VET students. Towards this end,we designed and implemented an Android-based app on Google Glass. Speaking time was chosen to give feedback based on existing literature. The tool was evaluated in a study consisting of a newly collected corpus of 15 students from local VET school in a dyadic sales pitch scenario.

The evaluation of questionnaire data provided insights about usefulness and distraction of the tool during a social interaction. An interesting observation has been the positive acceptance of glass by this age group in contrast to poor acceptance of GG in general. This could be due to novelty of the device or the fact that this generation may be more accepting of new technologies such as GG and similar devices as they are exposed to technology from an younger age. We believe this would be an interesting area to be explored in future. We also plan to explore the use of multiple nonverbal features for feedback including eye gaze or number of pauses. The challenges are to sense and process data by offloading intensive computation to a phone, without hindering dyadic interaction.

ACKNOWLEDGMENTS

This work was funded by the SNSF through the UBImpressed project. We thank Tanzeem Choudhury (Cornell Univ.), Marianne Schmid Mast & Denise Frauendorfer (Univ. Lausanne) for discussions, Michel Salamin (ECCG Martigny) for his support, and all student volunteers.

REFERENCES

1. Khaled Bachour, Frédéric Kaplan, and Pierre Dillenbourg. 2008. Reflect: An interactive table for regulating face-to-face collaborative learning. In *European Conference on Technology Enhanced Learning*. Springer.
2. Sumit Basu. 2003. A linked-HMM model for robust voicing and speech detection. In *Proc. of ICASSP*, Vol. 1. IEEE.
3. Steven A Beebe, Susan J Beebe, Mark V Redmond, and others. 2009. *Interpersonal communication*. Pearson.
4. Jared R Curhan and Alex Pentland. 2007. Thin slices of negotiation: predicting outcomes from conversational dynamics within the first 5 minutes. *J. Applied Psychology* 92, 3 (2007).
5. Ionut Damian, Chiew Seng Sean Tan, Tobias Baur, Johannes Schöning, Kris Luyten, and Elisabeth André. 2015. Augmenting social interactions: Realtime behavioural feedback using social signal processing techniques. In *Proc. ACM CHI*.
6. Joan Morris DiMicco, Anna Pandolfo, and Walter Bender. 2004. Influencing group participation with a shared display. In *Proc. of ACM Conf. on Computer supported cooperative work*. ACM.
7. Daniel Gatica-Perez. 2009. Automatic nonverbal analysis of social interaction in small groups: A review. *Image and Vision Computing* 27, 12 (2009).
8. Samuel D Gosling, Peter J Rentfrow, and William B Swann. 2003. A very brief measure of the Big-Five personality domains. *J. of Research in Personality* 37, 6 (2003).
9. Kiryong Ha, Zhuo Chen, Wenlu Hu, Wolfgang Richter, Padmanabhan Pillai, and Mahadev Satyanarayanan. 2014. Towards wearable cognitive assistance. In *Proc. of Int. Conf. on Mobile Systems, Applications, and Services*. ACM.
10. James G Hollandsworth, Richard Kazelskis, Joanne Stevens, and Mary Edith Dressel. 1979. Relative contributions of verbal, articulative, and nonverbal communication to employment decisions in the job interview setting. *J. Personnel Psychology* 32, 2 (1979).
11. Hayley Hung and Daniel Gatica-Perez. 2010. Estimating cohesion in small groups using audio-visual nonverbal behavior. *IEEE Trans. on Multimedia* 12, 6 (2010).
12. Taemie Kim, Agnes Chang, Lindsey Holland, and Alex Sandy Pentland. 2008. Meeting mediator: enhancing group collaboration with sociometric feedback. In *CHI'08 Extended Abstracts on Human Factors in Computing Systems*. ACM.
13. Mark Knapp, Judith Hall, and Terrence Horgan. 2013. *Nonverbal communication in human interaction*. Cengage Learning.
14. Hong Lu, Denise Frauendorfer, Mashfiqui Rabbi, Marianne Schmid Mast, Gokul T Chittaranjan, Andrew T Campbell, Daniel Gatica-Perez, and Tanzeem Choudhury. 2012. Stresssense: Detecting stress in unconstrained acoustic environments using smartphones. In *Proc. ACM UBICOMP*.
15. Gerard McAtamney and Caroline Parker. 2006. An examination of the effects of a wearable display on informal face-to-face communication. In *Proc. of SIGCHI Conf. on Human Factors in Computing Systems*.
16. Nathan Miczo, Chris Segrin, and Lisa E Allspach. 2001. Relationship between nonverbal sensitivity, encoding, and relational satisfaction. *Communication Reports* 14, 1 (2001).
17. Iftekhar Naim, M Iftekhar Tanveer, Daniel Gildea, and Mohammed Ehsan Hoque. 2015. Automated prediction and analysis of job interview performance: The role of what you say and how you say it. *Proc. IEEE FG* (2015).
18. Laurent Son Nguyen, Denise Frauendorfer, Marianne Schmid Mast, and Daniel Gatica-Perez. 2014. Hire me: Computational inference of hirability in employment interviews based on nonverbal behavior. *IEEE Trans. on Multimedia* 16, 4 (2014).
19. Eyal Ofek, Shamsi T Iqbal, and Karin Strauss. 2013. Reducing disruption from subtle information delivery during a conversation: mode and bandwidth investigation. In *Proc. ACM CHI*.
20. Harold Pashler. 1994. Dual-task interference in simple tasks: data and theory. *Psychological bulletin* 116, 2 (1994).
21. Alex Pentland and Tracy Heibeck. 2010. *Honest signals: how they shape our world*. MIT press.
22. Dairazalia Sanchez-Cortes, Oya Aran, Marianne Schmid Mast, and Daniel Gatica-Perez. 2012. A nonverbal behavior approach to identify emergent leaders in small groups. *IEEE Trans. on Multimedia* 14, 3 (2012).
23. Patrick E Shrout and Joseph L Fleiss. 1979. Intraclass correlations: uses in assessing rater reliability. *Psychological bulletin* 86, 2 (1979).
24. Janienke Sturm, Olga Houben-van Herwijnen, Anke Eyck, and Jacques Terken. 2007. Influencing social dynamics in meetings through a peripheral display. In *Proc of Int. Conf. on Multimodal Interfaces*. ACM.
25. M Iftekhar Tanveer, Emy Lin, and Mohammed Ehsan Hoque. 2015. Rhema: A Real-Time In-Situ Intelligent Interface to Help People with Public Speaking. In *Proc. of Int Conf. on Intelligent User Interfaces*. ACM.
26. Jens Weppner, Michael Hirth, Jochen Kuhn, and Paul Lukowicz. 2014. Physics education with Google Glass gPhysics experiment app. In *Proc. of ACM Int. Joint Conf. on Pervasive and Ubiquitous Computing: Adjunct Publication*.