Rhema: A Real-Time In-Situ Intelligent Interface to Help People with Public Speaking

M. Iftekhar Tanveer
ROC HCI, Electrical and Computer Eng.
University of Rochester
itanveer@cs.rochester.edu

Emy Lin
ROC HCI, Computer Science
University of Rochester
elin9@u.rochester.edu

Mohammed (Ehsan) Hoque
ROC HCI, Computer Science
University of Rochester
mehoque@cs.rochester.edu

ABSTRACT
A large number of people rate public speaking as their top fear. What if these individuals were given an intelligent interface that provides live feedback on their speaking skills? In this paper, we present Rhema, an intelligent user interface for Google Glass to help people with public speaking. The interface automatically detects the speaker's volume and speaking rate in real time and provides feedback during the actual delivery of speech. While designing the interface, we experimented with two different strategies of information delivery: 1) Continuous streams of information, and 2) Sparse delivery of recommendation. We evaluated our interface with 30 native English speakers. Each participant presented three speeches (avg. duration 3 minutes) with 2 different feedback strategies (continuous, sparse) and a baseline (no feeback) in a random order. The participants were significantly more pleased ($p < 0.05$) with their speech while using the sparse feedback strategy over the continuous one and no feedback.

Author Keywords
Public Speaking; Google Glass; Live Feedback; Design Techniques; Human Factor; User Study.

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION
Humans intuitively understand the elements of a well-presented speech. In addition to presenting meaningful content, the speaker needs to modulate his or her volume and vary speaking rate to retain the audience's attention [11]. Yet many presenters forget to do this in the delivery. During the act of publicly speaking, the speaker becomes the center of attention. In these kinds of scenarios, many people feel afraid and conscious of judgment, which often results in an overwhelming, uncomfortable, and stressful speaking experience. As a result, people rate public speaking as their number one fear (higher than the fear of death) [24]. With the advent of new and comfortable wearable technologies (e.g. Google Glass) and smart interfaces, a whole new realm of opportunities have arose to enable users to increase awareness of their nonverbal behavior during public speaking.

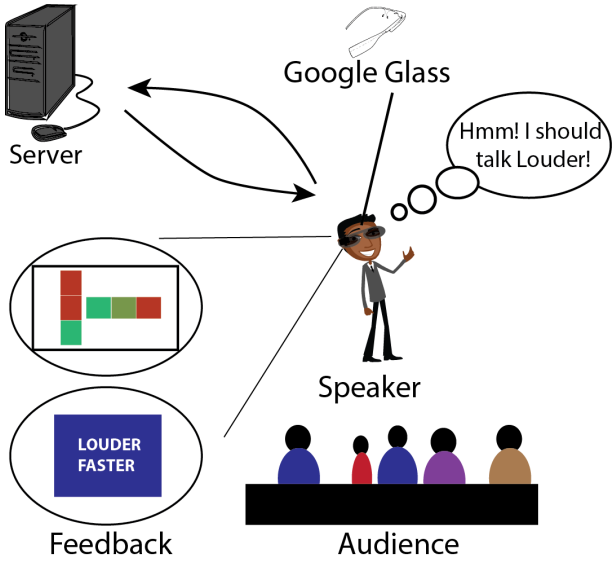

Figure 1: Usage scenario for Google Glass based real-time feedback system

In this paper, we present the design, development and evaluation of a smart user interface, Rhema[1], which presents live feedback on users' speaking styles through a wearable Google Glass. We have implemented a framework to record the live speech of the speaker using the Glass, transmit the audio to a server for automated processing of volume and speaking rate, and then present the data to the user. The analysis occurs in real time, allowing the interface to function during an actual speech delivery. The data is presented to the user in a format that is intuitive and informative about his or her voice modulation, without distracting them from delivering the speech and engaging with the audience.

IUI 2015, March 29 - April 01 2015, Atlanta, GA, USA
Copyright 2015 ACM 978-1-4503-3306-1/15/03…$15.00
http://dx.doi.org/10.1145/2678025.2701386

[1] www.cs.rochester.edu/hci/currentprojects.php?proj=rh

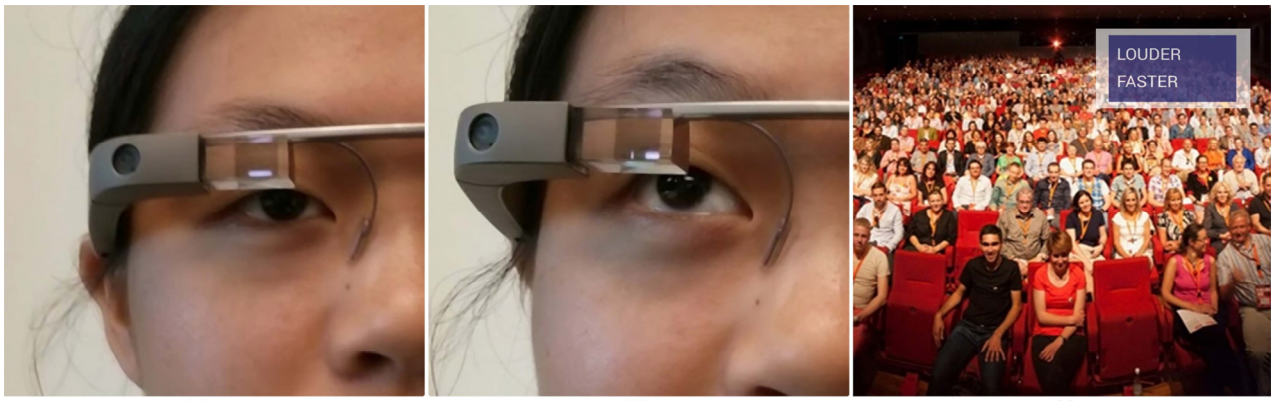

Figure 2: Demonstration of Google Glass as a secondary display (a) Eye position while looking forward (b) Eye position while looking at the glass display (c) User's view

Providing feedback during a live speech has a number of design challenges. One challenge is to keep the speakers informed about their speaking performance without distracting them from their speech. A significant enough distraction can introduce unnatural behaviors, such as stuttering or awkward pausing. Because the human brain is not particularly adept at multitasking [19,22], this is a significant issue to address in our feedback design. Secondly, the head mounted display is positioned near to the eye, which might cause inadvertent attention shifts [2]. Human attention is a limited resource [20], and the user might completely ignore the interface if the interaction requires too much attention. Additionally, if the user continuously stares at the feedback screen, the speaker will lose eye contact with the audience, causing the quality of the presentation to disintegrate.

We address these concerns by designing the interface in the form of "*Secondary Display*" [16], which avoids cognitive overload from multitasking and significantly improves the quality of the presentation. Information presented in secondary display is perceived and interpreted through a quick glance or a series of short glances. Unlike usual computer displays, it does not require a lengthy period of the users' attention. As a result, information presented in this form is less intrusive and minimizes the level of distraction. Physically, Google Glass works well as a secondary type because it is positioned in the periphery of the user's field of vision (Figure 2). The user must look upwards to actually see the display. In effect, it enables the users to usefully "*tradeoff attention for utility*" [16]. They make a glance towards the display if needed but do not lose attention inadvertently. A low requirement in attention makes the task cognitively less demanding, which results in more effective delivery of information. This is reflected in the results of our quantitative user study, in which the participants rated themselves significantly more satisfied ($p < 0.05$) with their speech while using the sparse feedback strategy than with the continuous one. They also strongly appreciated the minimization of distraction in comparison to what they had expected.

We came to an effective design decision by involving the potential users early and often in the design process. Starting with the brainstorming session, we reached out to a potential user group via social media and online forums. This enabled us to evaluate wide varieties of possibilities for the feedback design. We identified two main approaches for information delivery. The first approach consisted of continuous and slowly changing feedback schemes. This type of interface was designed to constantly deliver information through various plots, graphs or icons. To minimize distractions arising from movements in peripheral vision, this approach required the interface to change slowly [8]. The second approach was designed to deliver information sparsely over time. We designed a feedback scheme named "Words" which falls under this category. This feedback scheme shows nothing for 20 seconds. After 20 seconds, words such as "Louder", "Faster", "Good" (depending on the speaker's volume and/or speaking rate) are displayed. The words are shown for three seconds and then the display is blank again for the next 20 seconds.

We conducted a within-subject study with these two different types of feedback schemes along with a baseline case where Google Glass was turned off. 30 native speakers of English participated in this experiment. We collected qualitative and quantitative data from the participants. In addition, we collected quantitative ratings from 10 different Mechanical Turk workers for each speech to gain a sense of audience perception. The experiment demonstrated that the subjects found the system effective for adjusting their speaking rate. They also rated the sparse version of the feedback scheme significantly higher than the continuous version.

In summary, we contribute the following in this paper:

- We present a Google Glass interface to help people with public speaking by providing live feedback during the

delivery of the speech. This information enables the speakers to modulate their volume and speaking rate.

- We demonstrate that providing quick recommendations in regular intervals during the live speech delivery is more useful than continuously presenting the data to the users.
- Our experiment with 30 participants shows that the Google Glass interface adds value during the actual speech delivery in regards to voice modulation and speaking rate.

RELATED LITERATURE

Designing an interactive system that can capture and analyze a live speech and provide feedback using a wearable device draws on areas from Wearable Computing, Human Factors, Human-Computer Interaction, and Virtual Reality. Teeters et al. from MIT Media Lab demonstrated the use of self-cam [23], a custom built prototype where a camera hangs out from a person's neck and analyzes the facial expressions of the wearer. The prototype was developed with the need of individuals with autism to interpret their own emotions. This project allowed users to become self-aware by providing real time feedback on social behavior. While the prototype successfully demonstrated the feasibility of such technology, no user studies have been reported to measure its efficacy.

There is a growing body of literature in the area of Human Factors that focuses on developing ways to present live information to people while they are engaged in a cognitively overloading task. For example, Ofek et al. [18] studied the mode and the amount of information that can be consumed during a conversation. Their study suggested that it is possible to process information while a person is engaged in a conversation. However, people consume more information when the data is presented in small batches of visual elements. McAtamney and Parker reported that wearable technologies with active display (e.g. the display of handheld smart phones) significantly disrupt social interactions because users lose the element of eye-to-eye contact [15]. We kept these findings in mind while designing our interface, and they also guided our decision to use Google Glass as a secondary display instead of an active head-mounted display.

Researchers continue to study both verbal and nonverbal behaviors during public speaking in an effort to find ways to improve speaking skills. Eva et al. described a study in which they analyzed the effect of acoustic features of individuals perceived as excellent speakers [21]. They conducted a perceptual evaluation by manipulating the F0 dynamics, fluency, and speech rate in a synthesized animation of a speaker. They noted that the amount of manipulation is highly correlated with the ratings of the speaker. Koppensteiner et al. [13] reported the effect of "dominant activation of body parts" with different personality traits. However, scopes of all these previous efforts are limited to manually inspecting and understanding positive and negative behaviors of public speaking. There is room for further research by automating the analysis and providing feedback.

A considerable amount of effort has taken place related to public speaking in the virtual reality domain as well. In 1998, North et al. described a study in which virtual reality was used as a therapy to treat phobia associated with public speaking [17]. In the study, a group of people was exposed to a virtual public speaking scene. The group showed significant improvements after five weeks of treatment. This experiment provided evidence in support of virtual reality therapy for improving public speaking skills.

Researchers have also worked on virtual environments to enable people to practice public speaking. Chollet et al. from University of Southern California developed an interactive virtual audience program for public speaking training [5]. The program collects a dataset of public speaking performances under different training conditions. The same team of researchers is currently working to build a fully multimodal platform named "Cicero" [1] to automatically analyze and train a user's behavior during public speaking. Cicero can identify some nonverbal descriptors that are considered characteristics of speakers' performances. However, providing any user level feedback to improve speaking performances remains part of their future work. Hoque et al. developed an intelligent avatar coach named MACH, "My Automated Conversation coacH" [12], to help people practice job interviews. Their results suggested that a virtual coach might be more effective than traditional behavioral interventions. However, all the aforementioned applications only allow participants to practice social interactions offline. In this paper, we take on the challenge of providing live feedback during the actual interaction.

TECHNICAL DETAILS

In order to validate whether real time feedback would actually elicit measurable improvements, we implemented working prototypes on Google Glass using the Android platform. We chose Google Glass as it is lightweight and comfortable to wear and comparatively cheaper than other head mounted displays. We hoped that people using this form factor would not become overwhelmed from the discomfort of a traditional head-mounted display. In addition, as shown in Figure 2, the position of Google Glass display worked effectively for communicating information in real time without inadvertently distracting the user. Unfortunately, Google Glass is not designed to run computationally intensive programs [10]. It does not have any suitable heat dissipation mechanism. As a result, running intensive algorithms to process the speakers' speeches made the glass hot and uncomfortable for long continuous use.

To overcome this problem, we used a local server for leveraging the computational capabilities. The server

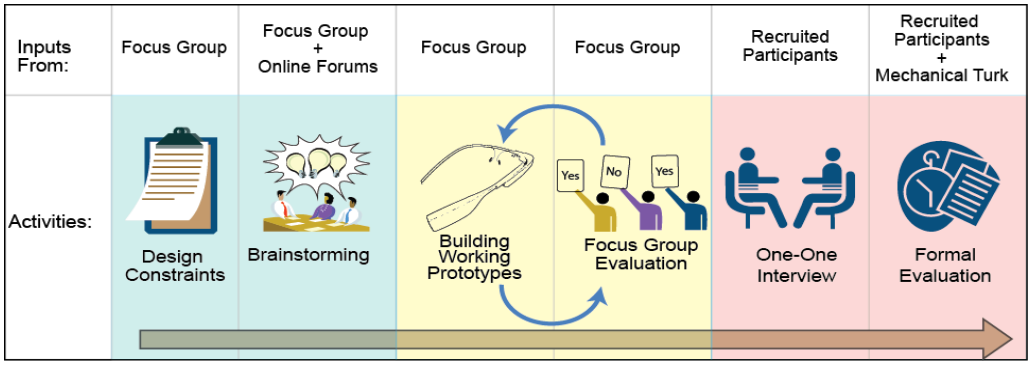

Figure 3: Chronological list of actions taken for finding a suitable design of a live feedback scheme

analyzed the audio data for loudness (in decibels) and speech rate (in words per second) and sent the results back to the glass. We experimented with Transfer Control Protocol (TCP), Real-time Transport Protocol (RTP) and User Datagram Protocol (UDP) to transmit each second of audio clips to the local server. The use of streaming protocols like RTP caused the Glass to overheat. On the other hand, UDP datagrams were too unreliable for accurate calculation of loudness and speaking rate as there is no guarantee for packet delivery, ordering and duplicate protection. Therefore, TCP was suitable for our application as the audio clips were only used in calculation of loudness and speaking speed and not for playing them in real-time. Offloading the computationally expensive tasks to a local server enabled us to use the Google Glass for about an hour without overheating it. However, with further upgrade to the hardware, it may be possible to extend the battery life in the future.

On the server-side, we calculated the perceived loudness from the audio signal energy. Because we recorded the audio signal from the microphone embedded in Google Glass, the recorder settings remained consistent for each participant. As a result, calibrating the intensity levels once for a single user gives us highly accurate loudness estimates for all of the participants. We used an open source audio processing library named PRAAT [4] to calculate the loudness levels for the participants. In order to run experiments to validate the interface, we calibrated the app based on the audio settings in our lab, where the ambient loudness was 30dB. As a result, we empirically determined that a range of 54dB to 58dB to be a medium range for giving a speech. Less than 54dB was regarded as "quiet" and greater than 58dB as "loud". We empirically adjusted these ranges and remained constant throughout the formal evaluation period. However, the application has the functionality to automatically calibrate these thresholds levels for each individual's natural level of loudness, allowing it to be deployed outside of the lab.

We employed an estimation of fundamental frequency to detect the voiced and unvoiced regions in the audio signal. From these regions we calculated the speaking rate of the participants. We used the pitch detection algorithm available in PRAAT proposed by Paul Boersma [3]. We created a routine to count the number of discontinuities within the pitch contour. This count gave a rough estimate of the number of words per second. This was a simple metric and fast enough for our application. The server stored the loudness and speaking rate values of the last five seconds to calculate a weighted average of the samples. We used appropriate weights to resemble the effect of a "leaky integrator" so that it removed the effect of high frequency noise from the calculated values of speaking rate.

METHODS

One of the fundamental challenges of this research was to design a feedback interface that can convey loudness and speaking rate measurements to a speaker with minimal distraction. To achieve this goal, we performed a principled design strategy as illustrated in Figure 3. We involved our users early on in the design process. In this regard, we formed a focus group for quick evaluation and redesign of the prototypes. The participants of the focus group were students and faculty members who regularly met during weekly Human-Computer Interaction (HCI) meetings at the department of Computer Science over the summer of 2014. Members of this focus group included two professors in Computer Science, two PhD students, one research programmer, and five undergraduate students. We describe each activity in the following sections.

Design Constraints

Based on our theoretical understanding, we decided that the intended feedback interface should use a secondary form of information delivery to avoid cognitive overload. Additionally, the feedback scheme should contain as few movements as possible in order to avoid unintentional attention drift. Lastly, the interface should contain temporal history of the user's performance. As we wanted to encourage the users to modulate their voice, showing only an instantaneous value of loudness and speaking rate was not enough to understand how they were modulating their voice.

Brainstorming Session

Before brainstorming the possible ideas and design choices for the interface, we solicited ideas from Reddit.com in the

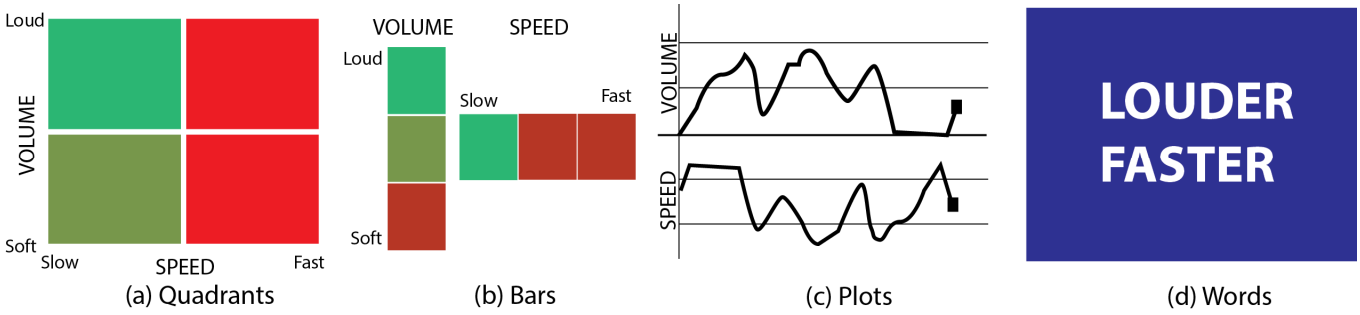

Figure 4: A few Feedback schemes that we explored for our interface

form of a survey. We particularly chose the forums "Futurama" and "HackerNews" since these communities are open to new and innovative ideas. We obtained a mixed response in this survey. 4 people out of 28 total respondents mentioned that it would be a bad idea to provide any information while a person is speaking because it would severely distract the speaker. However, many others proposed several feedback schemes. Seven respondents proposed displaying only raw numbers in decibels and words per second as a feedback scheme. Five participants proposed various forms of bar graphs (horizontal for speed, vertical for loudness, bars changing colors, decibel meters, etc.). Others proposed various icons, line graphs or traffic colors to represent volume and speaking rate.

In the brainstorming session with the focus group, we reviewed the proposed feedback schemes. Everyone in the focus group agreed that raw values or simple bar graphs would not be sufficient according to the design constraints because they do not contain enough temporal history. We converged on a few potentially viable designs from the brainstorming session as shown in Figure 4. The "*Quadrant"* feedback scheme (Figure 4a) resembles a 2D graph. The volume is on the y-axis and speed is on the x-axis. This graph is divided into four quadrants, each representing a range in volume and speed. When the user's voice is within a certain range, the corresponding box fades to green. If the user's voice hits a different range, that range will begin to fade to green and the previous range will start to fade back to red. The fading is slow enough so that more than one box can be green at once. Ideally, all the boxes turn green during a speech, indicating an appropriate amount of voice modulation. Presence of red boxes indicates that the user is not hitting that range for a considerable amount of time. The "*Bars*" feedback scheme (Figure 4b) involves two bars, each segmented into three regions. The vertical bar represents volume and the horizontal one is speed. The fading between red and green works the same way as in the Quadrant feedback scheme. Orientations of the bars were decided from people's natural tendency to associate high/low volume with up/down and fast/slow speed with left/right. This design allowed the users to immediately identify in a quick glance which bar is for volume and which one is for speaking rate. These two bars change independently of one another. The "*Plot"* feedback scheme (Figure 4c) plots raw data on two separate line graphs, one of volume on the top half of the display and one of speed on the bottom half. The Plot Feedback Scheme was selected for its ability to present temporal information and its intuitive representation of data, thus eliminating the learning curve. The "*Words*" feedback scheme (Figure 4d) displays nothing for 20 seconds; during this period, the system measures the speaker's voice modulation. For example, if the speaker spoke too quietly for a significant portion of the 20 seconds, then the word "LOUDER" will be displayed on the screen. Similarly, if the speaker spoke too loudly during the 20 seconds, the word "SOFTER" will appear on the screen. The same goes for speed ("FASTER" or "SLOWER"). If the speaker has adequately modulated their voice, either in volume or speaking rate or both, "good!" is displayed. These words are displayed for three seconds and then the Glass display becomes blank again for another 20 seconds. The 20 seconds of time interval was decided in our iterative informal evaluations. We found that any additional increase in this duration causes the user to question whether the app is working.

Iterative Implementation and Evaluation of Prototypes

In the next step, we started an iterative implementation and evaluation cycle. We implemented different versions of the Quads, Bars, Plots, and Words feedback scheme and evaluated empirically in the focus group discussions. We also built a number of other innovative feedback schemes. For example, we implemented "Audio Feedback" which functions like a metronome. It produces a ticking noise at a rate determined by the user's speaking rate. When the user talks too quickly for too long, the system starts ticking at a slower pace to encourage the user to slow down, and vice versa. Another feedback method we implemented is "Black and White". The display fades between black and white depending on the loudness of the user's voice. However, the last two prototypes could deliver only one variable (either volume or loudness) at a time. Using both schemes together was found to be too distracting and overwhelming.

One-to-One Interview Session

As we wanted to select only two feedback schemes for the formal evaluation, we needed to rank among the prototypes in terms of their efficacy. We arranged one-to-one

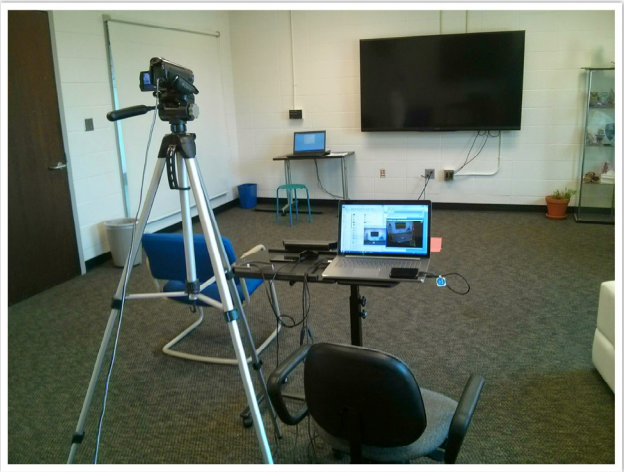

Figure 5: The setup for formal evaluation

interview sessions with 13 students to rank these prototypes. Before the interview session, all the participants were unfamiliar with the project. In total, 9 female and 4 male students participated. The participants ranged from 19 to 25 years old. We presented them with the Google Glass prototypes, explained the goal of the project, and allowed each participant to try each of the feedback schemes while reciting a poem or talking about him or herself. Once they were accustomed with all the prototypes, we asked them to choose their top two preferred versions of the interface and explain their choice. From their feedback, we found the

Statements related to speech	Overall, I am happy with the quality of my speech; I varied my volume appropriately; My speech was well paced; I showed appropriate body language; I maintained eye contact with audience
Statements related to Feedback Scheme	**Efficacy:** I found this feedback to be very helpful; The quality of my speech improved due to the feedback; The feedback was useful in helping me modulate my volume; The feedback was useful in helping me vary my speed; I felt distracted by the feedback; The usefulness of the feedback outweighed any distractions; **Learnability:** The feedback scheme was easy to learn; It was easy to follow while delivering my speech; The learning curve for using the feedback system is huge; **Future Use:** If available, I would love to use this feedback system in an actual speech

Table 1: List of statements as in the post-speech survey. The live form is available at http://tinyurl.com/rhemaSurvey

Figure 6: Sample snapshot from a recorded video of public speaking

following ranking: Words (11), Bars (8), Quadrant (6), and Plots (1). The quantitative data highlights the users' preference for sparse feedback over continuous feedback. This was further echoed in the participants' qualitative responses.

EVALUATION

We evaluated Rhema in order to seek answers to the following questions:

- What value, if any, is added to the participants' performance while using the interface?
- Do the participants find the interface effective and easy to learn?
- Are there any noticeable side effects or distractions from the usage of Rhema?

Procedure

In order to answer these questions, we designed a formal user study where we asked the participants to deliver three speeches, three minutes in duration, while wearing Google Glass. Two speeches were delivered with the Google Glass display on, one with the "Words" feedback scheme and the other with the "Bars" feedback scheme. Another speech was delivered with the Google Glass display turned off, which served as a baseline. The participants wore Google Glass for the baseline to ensure that the results would not be influenced due to the mere existence of the Glass. Before the participants delivered the speeches with the Bars and Words feedback schemes, we explained the use of the interface to the users and set aside at least five minutes so that they may practice and familiarize themselves with it. We began the actual speech only when the participants said they were comfortable with the feedback scheme. The speakers chose three topics from a list of sample topics (e.g. favorite pastime, favorite book/movie/superhero, etc.) that we supplied for convenience. The topics were decided at least two days ahead of the presentation time to allow the participants to prepare. The order of both the topics and the Google Glass feedback schemes were counterbalanced to remove any ordering effects.

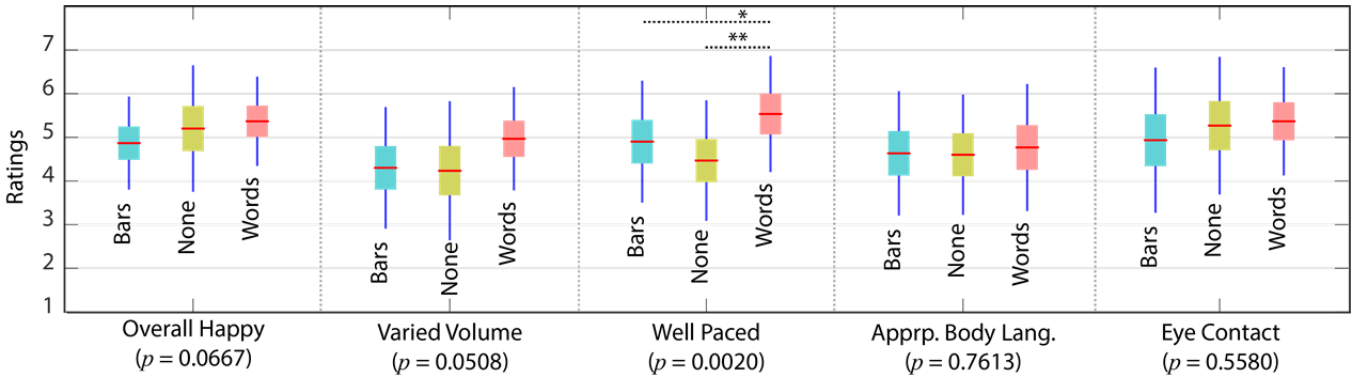

Figure 7: Boxplot of responses for each measure enlisted in the first row of Table 1

The study was conducted in a lab environment where each speech was recorded using a high definition video camera, as shown in Figure 5. Figure 6 shows a sample snapshot of the videos that we recorded. The camera was placed to capture the participant and a part of the audience's head. This allowed the viewers of the videos to judge if the speaker was maintaining eye contact with the audience. A brief interview session regarding the participants' experiences with the feedback schemes took place at the end of each study session.

Participants

We recruited 30 students from University of Rochester, of which 17 were male and 13 female. Their ages ranged from 18 to 32 and average age was 20. There were 10 freshmen, 4 sophomores, 4 juniors, 5 seniors, 4 graduate students, and 3 others. We posted flyers at different places throughout the campus, and we also posted recruitment messages in different Facebook groups associated with student activities. The study was limited to native speakers of English of at least 18 years of age.

Measures

To assess the extent of value that Rhema added to the participants' experiences, we asked our participants to fill out a survey related to their speaking performance. Participants filled it out immediately after delivering each speech. These measures are listed in the first row of Table 1 and are related to the participant's overall speech.

We asked the participants to fill out a different set of measures to assess the efficacy, learnability, and future use of the interface (Listed in the second row of Table 1). All the measures were answered in a 7-point Likert scale where 7 represents 'strongly agree.' Since these measures are only related to a particular feedback scheme (e.g. easy to follow, the associated learning curve, etc.), these are not applicable when no feedback exists. Thus, we have responses in these measures for Bars and Words feedback only.

Finally, to assess any noticeable effects of distractions arising from Rhema, we posted each video to Amazon Mechanical Turk. Ten different workers evaluated the videos. Table 2 presents the measures of ratings. The measures were selected to identify any possible effect of distraction. In order to ensure the quality of Mechanical Turk ratings, we opened the assignments only to workers with a good working history (i.e. master workers with >95% HIT acceptance rate and a total of >5000 accepted HITs). We considered the ratings from Mechanical Turk workers as anonymous opinions from a general audience.

RESULTS

Figure 7 illustrates a boxplot representing the participants' responses to the questions related to their speech. The responses are shown under three different groups representing three different feedback schemes. In the boxplot, the red horizontal line represents sample mean, the blue vertical line represents 1 standard deviation range and the box represents 95% confidence interval in t-test (for illustration purpose only). As the users rated on an ordinal scale (7-point Likert), we used non-parametric methods of calculating statistical significance. We used Friedman's test [9] to detect the differences among the groups. The *p*-value for this test is given within parentheses along the horizontal axis. Upon finding any significant differences among the groups, we used Wilcoxon's Signed Rank [25] test for pairwise comparison. We applied Bonferroni correction [7] to counteract the problems of multiple comparisons. In the box plot, we used one or two asterisks to represent differences with significance level 0.05 and 0.01

Statements	Kα
This person did not pause too often during the speech	0.28
This person maintained eye contact with whomever he or she was talking to	0.35
This person did not use a lot of filler words (Um, uh, basically etc.) during the speech	0.39
This person does not appear distracted	0.22
This person did not appear stiff during the speech (i.e. looks spontaneous)	0.30

Table 2: List of statements in the Mechanical Turk worker's questionnaire. Kα represents the inter-rater agreement using Krippendorff's Alpha [14].

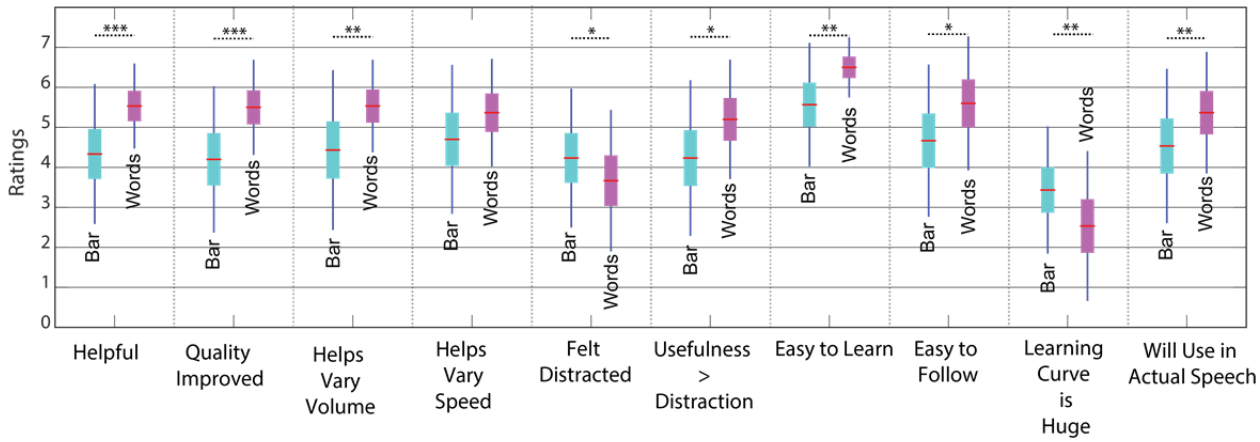

Figure 8: Boxplot of responses for each statement listed in the second row of Table 1

respectively. The plot illustrates that the participants rated their speech to be significantly well paced while using the Words feedback scheme. The Friedman's test rejects the null hypothesis with a *p*-value of 0.002 for this case. The *p*-value for the measure on whether the participants varied their volume was slightly higher than the significance level and thus failed to reject the null hypothesis.

Figure 8 represents the boxplots of responses to the measures related to the interface. As there are only two groups to be compared in this plot, we used Wilcoxon's Signed Rank test [25] to calculate the statistical significance. In the figure, we used one, two, or three asterisks to represent 0.05, 0.01, and 0.001 significance level. The plot shows that the Words feedback scheme is rated significantly better in most of the survey measures.

Measures from Mechanical Turkers

We considered the ratings provided by the Mechanical Turk workers in order to assess viewer's opinion (10 viewers per participant). While it is easy to get Mechanical Turk ratings, the workers are less likely to be experts in public speaking. Therefore, it was important to assess the reliability of the ratings. To determine the quality of the ratings, we calculated inter-rater agreement using Krippendorff's Alpha [14] as shown in Table 2. Kippendorff's alpha is a better metric than other methods of calculating inter-rater agreement (e.g. Cohen's Kappa [6]) because it allows any number of raters, missing data, and any type of measurement values (e.g. binary, nominal, ordinal, interval, etc.).

In our results, the agreement varied from 0.229 to 0.395 for different survey measures. We noticed that the agreement was highest for objective questions that asked the Turkers to rate the quality of eye contact and the use of filler words. On the other hand, it was lowest for the question on the speakers' level of distraction. We suspect different people might consider different criteria for assessing distraction (e.g. it might mean not making eye contact, not moving, or repeating the same words), which may have led to this poor agreement between the raters in this measure. Figure 9 shows the box plot of the responses.

To calculate the statistically significant differences in the ratings under different feedback schemes, we conducted Friedman's test. The *p*-values are reported within parentheses in Figure 9. Unfortunately, the test failed to reject the null hypothesis for any of the measures. This may result from the fact that the Turkers are not as trained as public speaking experts to identify nuance differences in a public speaker's performance. The suboptimal agreement in Turker's ratings also indicate that possibility.

Post Study Interview Results

As evidenced in the brief post-study interview sessions with the formal evaluation of participants, some form of real-time feedback during a live speech certainly has a beneficial effect. We asked each participant, "*Was x feedback useful or effective in reminding you to modulate*

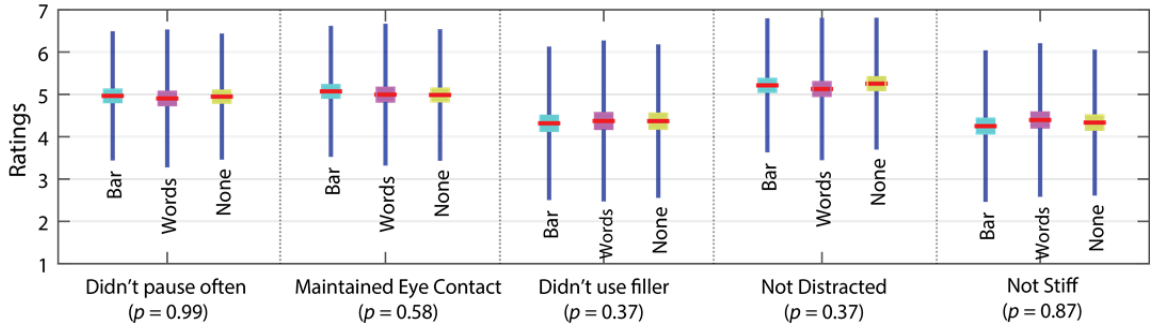

Figure 9: Boxplot of responses from Mechanical Turk Workers

your voice?" All replied yes to either one or both of the feedback schemes they tried. Several reflected on their three speeches and concluded that the feedback "*definitely helped*" and "*made [them] more aware*" of their voice modulation, especially when they compared their speech with feedback to that with no feedback. Regardless of the type of feedback scheme, all participants said that the customized feedback "*actively changed how I spoke more*" or that "*it reminded me that I was supposed to change something*".

We also asked the participants about their preferred feedback scheme. 22 of the 30 participants preferred the Words feedback scheme, rather than the Bars, because the Words feedback was straightforward and simpler. Those who did not like Words complained that the feedback required additional effort from the language part of the brain since they were already using it for the content of their speech. However, other participants disagreed. One stated that the Words feedback scheme was "*less cognitively demanding while giving a speech*" as it was "*more straightforward*." Several found it useful that the Words feedback scheme displayed "*information that could be parsed a little more quickly*" as it only showed two pieces of data for a few seconds at a time.

> *"It was just enough information that I could take in while also talking, but not too much information that it was overwhelming."*

Eight of the 30 participants preferred the Bars feedback scheme rather than the Words. Some preferred it because it contained more information that was always readily available on the screen, and this constant feedback felt reassuring to them. They also liked the visual aspect of it and that the "*visual nature of it didn't interfere with the words [they] were thinking of*." However, many did not find the Bars useful or effective. Several participants did not know how to interpret or apply the information to their voice. The Bars feedback scheme was also clearly confusing; two participants explicitly told us during the post-speech interview that they wondered, "*What am I supposed to do with this?*" and "*What does it mean?*" while using the Bars feedback. Some found it too overwhelming and complex and ended up completely ignoring it. One participant explained that

> *"It was so complicated that I couldn't use it. The pause I would have to take to process the information was so great that I couldn't maintain my speech."*

Aside from these general overall trends in opinions on the feedback schemes, personal preferences drove many of the participants' responses. For example, some are more receptive to visuals – those who preferred the Bars feedback scheme thought it worked well because they consider themselves "*visual thinker[s]*". They also varied in the amount of information they wanted to see: one participant suggested using less boxes in the Bars feedback system, and another suggested using many more boxes; some wanted to see the words appear more frequently, whereas others thought the implemented frequency was good. Some participants thought the Bars feedback was easier to ignore, while others were more likely to disregard the Words feedback. From this, we concluded that the efficacy, usefulness, and level of distraction of the different feedback schemes vary largely as a result of the users' personal preferences and tendencies.

DISCUSSIONS AND CONCLUSION

In this work, we implemented a fully functional, real time Google Glass interface to help people with public speaking. The interface provides live feedback regarding the user's volume and speaking rate. The primary objective was to maximize the usefulness of our system by minimizing the level of distraction.

From the quantitative and qualitative evaluations of our system, we can safely conclude that the participants found the interface valuable to their speaking performance. The qualitative study reflects that the participants rated the Words feedback to be effective in varying their speaking rate and significantly easier to learn than the Bars feedback scheme. We hypothesize that this is the effect of our Sparse information delivery strategy and recommendation-based feedback scheme. Unfortunately, we could not conclusively measure the level of distraction from the audience's perception. As the Mechanical Turk raters were not experts in public speaking, their ratings did not agree and also failed to pass any hypothesis tests. It is possible to address this in the future by involving expert opinions from the Toastmasters International Organization.

We collected Kinect [26] depth information while the participants delivered their speech, and we will leverage this dataset to train machine learning algorithms that can analyze and predict speaking performance from a combination of the speaker's voice patterns, facial expressions, and body language. We will show these predicted performances to the users and offer live recommendations to help them further improve their public speaking skills and speech delivery.

This research also has a number of implications for future endeavors. Presenting information sporadically through secondary display might be a generic strategy in situations with cognitive or attentional overload. The preliminary observation described in this paper calls for further investigation in this area. In addition, we see room to explore the potential for haptic feedback schemes in similar situations.

Our current prototype was originally designed to give novice speakers real-time feedback on voice modulation. However, as the system already sends the audio data to a server, this data could be used to provide post-speech feedback to the user. Such a system could be useful for even expert users because it would allow them to review the

speech and check for mistakes in retrospect. Without the need for instantaneous feedback, such a system could be adapted to a range of wearable technologies including smart phones. Rhema is available for download in the following webpage:
www.cs.rochester.edu/hci/currentprojects.php?proj=rh

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. *Intelligent Virtual Agents*, Springer (2013), 116–128.
2. Biocca, F., Owen, C., Tang, A., and Bohil, C. Attention Issues in Spatial Information Systems: Directing Mobile Users' Visual Attention Using Augmented Reality. *Journal of Management Information Systems 23*, 2007, 163–184.
3. Boersma, P. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *IFA Proceedings 17*, (1993), 97–110.
4. Boersma, Paul ;Weenink, D. Praat: Doing Phonetics by Computer. *Ear and Hearing 32*, 2011, 266.
5. Chollet, M. et al. An interactive virtual audience platform for public speaking training. *Autonomous agents and multi-agent systems*, (2014), 1657–1658.
6. Cohen, J. A coefficient of agreement of nominal scales. *Educational and Psychological Measurement 20*, (1960), 37–46.
7. Dunn, O.J. Multiple Comparisons Among Means. *Journal of the American Statistical Association 56*, (1961), 52–64.
8. Finlay, D. Motion perception in the peripheral visual field. *Perception 11*, 1982, 457–462.
9. Friedman, M. A comparison of alternative tests of significance for the problem of m rankings. *The Annals of Mathematical Statistics 11*, 1 (1940), 86–92.
10. Ha, K., Chen, Z., Hu, W., Richter, W., Pillai, P., and Satyanarayanan, M. Towards wearable cognitive assistance. *MobiSys '14*, (2014), 68–81.
11. Hoogterp, B. *Your Perfect Presentation: Speak in Front of Any Audience Anytime Anywhere and Never Be Nervous Again*. McGraw Hill Professional, 2014.
12. Hoque, M., Courgeon, M., and Martin, J. Mach: My automated conversation coach. *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*, (2013), 697–706.
13. Koppensteiner, M. and Grammer, K. Motion patterns in political speech and their influence on personality ratings. *Journal of Research in Personality 44*, (2010), 374–379.

ACKNOWLEDGEMENT

This research was developed through a generous gift from the Google Faculty Research Award. We would like to thank all the participants in the research study. We thank our anonymous reviewers for many helpful comments.

14. Krippendorff, K. *Content Analysis: An Introduction to Its Methodology*. 2004.
15. McAtamney, G. and Parker, C. An examination of the effects of a wearable display on informal face-to-face communication. *Proceedings of ACM CHI 2006 Conference on Human Factors in Computing Systems*, (2006), 45–54.
16. McCrickard, D.S., Catrambone, R., Chewar, C.M., and Stasko, J.T. Establishing tradeoffs that leverage attention for utility: Empirically evaluating information display in notification systems. *International Journal of Human Computer Studies 58*, (2003), 547–582.
17. North, M.M., North, S.M., Coble, J.R., et al. Virtual reality therapy: an effective treatment for the fear of public speaking. *The International Journal of Virtual Reality 3 (1998), 1 3*, (1998).
18. Ofek, E., Iqbal, S.T., and Strauss, K. Reducing disruption from subtle information delivery during a conversation: mode and bandwidth investigation. *Proceedings of CHI 2013*, (2013), 3111–3120.
19. Pashler, H. Dual-task interference in simple tasks: data and theory. *Psychological bulletin 116*, (1994).
20. Shiffrin, R.M. and Gardner, G.T. Visual processing capacity and attentional control. *Journal of experimental psychology 93*, (1972), 72–82.
21. Strangert, E. and Gustafson, J. What makes a good speaker? subject ratings, acoustic measurements and perceptual evaluations. *INTERSPEECH*, (2008), 1688–1691.
22. Strayer, D.L., Drews, F.A., and Crouch, D.J. Fatal distraction? A comparison of the cell-phone driver and the drunk driver. *Human Factors: The Journal of the Human Factors and Ergonomics Society 48*, (2006), 381–391.
23. Teeters, A., Kaliouby, R. El, and Picard, R. Self-Cam: feedback from what would be your social partner. *ACM SIGGRAPH 2006 Research posters*, (2006).
24. Wallechinsky, D. *The book of lists*. Canongate Books, 2009.
25. Wilcoxon, F. Individual comparisons of grouped data by ranking methods. *Journal of economic entomology 39*, (1946), 269.
26. Zhang, Z. Microsoft kinect sensor and its effect. *MultiMedia, IEEE 19*, 2 (2012), 4–10.