Reducing Disruption from Subtle Information Delivery during a Conversation: Mode and Bandwidth Investigation

Eyal Ofek, Shamsi T. Iqbal and Karin Strauss
Microsoft Research
One Microsoft Way, Redmond, WA 98052
{eyalofek, shamsi, kstrauss}@microsoft.com

ABSTRACT
With proliferation of mobile devices that provide ubiquitous access to information, the question arises of how distracting processing information in social settings can be, especially during face-to-face conversations. However, relevant information presented at opportune moments may help enhance conversation quality. In this paper, we study how much information users can consume during a conversation and what information delivery mode, via audio or visual aids, helps them effectively conceal the fact that they are receiving information. We observe that users can internalize more information while still disguising this fact the best when information is delivered visually in batches (multiple pieces of information at a time) and perform better on both dimensions if information is delivered while they are not speaking. Interestingly, participants qualitatively did not prefer this mode as being the easiest to use, preferring modes that displayed one piece of information at a time.

Author Keywords
Augmented Reality; Attention; Design; Human Factors.

ACM Classification
H.5.2 [Information interfaces and presentation]: User Interfaces. - Graphical user interfaces.

INTRODUCTION
The increasing ubiquity of mobile devices is making information available anywhere, anytime [19, 22, 25]. The development of mobile processing power, the inclusion of sensors such as GPS, accelerometers, gyros, cameras and microphones, as well as the accumulation of supporting data and computation in the cloud, made it easy to display information to a user based on location and context [33]. Easy and seamless access to information can help augment many ongoing tasks, e.g., providing awareness of peripheral activities [6, 34], opportunistically delivering information to help coding and development [14], and providing relevant information during searching [3, 13].

A major challenge when it comes to openly consuming information in social settings is the perceived cost in terms of disruption to established social norms of being attentive to the ongoing interaction [2, 10, 18, 27]. Widespread adoption of such practices largely depend on their social acceptability: how seamlessly they fit into the routine activities carried out in everyday life [30].

Our broader research goal is to investigate how to deliver information to augment natural conversation among people in face-to-face settings. Presenting conversational aids may improve conversation quality; e.g., 7-9 words delivered peripherally has been shown to help trainee supervisors better manage training sessions [26]. Such aids may benefit other common real world scenarios, e.g., short text or words may help strangers find topics of mutual interest, and help acquaintances recall the context of prior meetings. However, expected social norms may deter people from explicitly seeking for aid as processing information in the background during a conversation can be potentially distracting. Research on conversation aids has mostly focused on helping people with cognitive degenerative diseases such as dementia and Alzheimer's [11, 28], but for this population the goal is to help them participate in settings where they are unable to communicate independently and the disruption caused by the device is typically overlooked.

For more general settings, we wish to provide conversation aids that do not disrupt ongoing conversations. We focus on minimizing distraction, as prior work has shown explicit consumption of information during a conversation can be interpreted as a lack of interest [10]. As a first step towards this goal, we studied how information processing can be made more discreet and socially acceptable. More specifically, we wanted to understand how much information users can internalize without any discernible cues during a conversation, and how this was impacted by the medium and mode of information delivery. We conducted a user study that targeted measuring how much information users can consume during a conversation while not letting it show to their interlocutor. We exposed participants to three different modalities and multiple word group sizes with the purpose of better understanding how these factors affect their ability to maintain a conversation with sufficient attention while effectively consuming content.

Results showed that while participating in a conversation, users can internalize more information while still disguising this fact the best when information is visually delivered in batches (multiple pieces of information at a time). Addi-

CHI 2013, April 27–May 2, 2013, Paris, France.

tionally, users perform better on both dimensions if information delivery is done while they are not speaking. Interestingly, participants qualitatively did not prefer this mode as being the easiest to use, preferring modes that display one piece of information at a time. Our findings have important implications for the design of augmented reality systems intended to deliver information while the receiver performs other tasks, e.g., face-to-face conversations.

RELATED WORK

We discuss related work in peripheral awareness, and device use and social communication.

Devices for peripheral information delivery

Peripheral awareness systems allow multitasking users to access information without causing abrupt suspension of ongoing tasks. Maglio and Campbell [15] compared various modes of displaying peripheral information to a multitasking user to understand how the mode affects the balance between delivering information and distraction from the primary task. McCrickard et al. [18] explored ways of presenting peripheral information to users engaged in a browsing task. They found tickering text to be the most effective and least disruptive to the primary task, and that smaller displays were more disruptive, while slower presentation of information was more comprehensible. In the domain of driving, Brumby et al. compared the effectiveness of delivering information to drivers via audio and visual displays [5]. They found that participants who prioritized processing the information over driving safety preferred visual display because of the speed of processing, whereas participants who prioritized safety chose audio – suggesting that the sharing of cognitive processing channels while multitasking results in performance compromise on at least one of the tasks, even if the task is more suited for that channel.

In the domain of augmented reality (AR) devices, the practice of adopting AR technologies to enhance how we perceive and experience the world is rapidly becoming ubiquitous [32]. In such devices, information may be transmitted to the user in the form of computer generated sensory input such as sound or video. The proliferation of positional and directional sensors in smart phones enables the distribution of AR applications to the general public [20]. For example, the Remembrance Agent project [23] was aimed at delivering just-in-time information for users using a wearable display. The focus of this and later work was on recovery of relevant information and usability for a single user.

Over the years head-mounted displays (HMDs) and virtual retinal displays for visualization have been researched extensively [7, 16, 21, 29]. Unfortunately, many of these displays are cumbersome, and are non-natural in appearance, which limited their use to few early adopters. More recent HMDs, such as Google's project Glass [1], have the advantage of being lightweight and appear similar to accessories people already use – thus such displays have the potential of being more socially acceptable. However, current implementations are not quite transparent, and display information on the periphery, causing people to look away.

In context of the prior work on peripheral information systems, our work looks at the mode and bandwidth of information that can be delivered to a person without disrupting the ongoing flow of a face-to-face conversation. While the use of teleprompters is common for newscasters and reporters for receiving information surreptitiously, in these cases information delivery is mostly unidirectional. No formal study has investigated effectiveness of information delivery in bi-directional settings, e.g., conversations. The challenge is to determine the optimal bandwidth and appropriate mode that helps users easily consume information without giving out cues that they are receiving that information, so that social protocols of being attentive during face-to-face conversations are not violated.

Device use and social communications

The use of devices to access information in social settings is rapidly becoming a common practice, but the reaction to resulting distractions is not well understood. Iqbal et al. studied the use of devices during presentations and reported on the perceived costs of disruption on both speakers and device users, and benefits of information awareness for device users [12]. Campbell and Kwak found that use of technology in public did not detract people from conversing to strangers in public [8], suggesting that using appropriately designed technology during face-to-face conversation may blend in with such social interactions. McAtamney and Parker studied how wearable computing devices affect face-to-face conversation [17]. Their results showed that wearable devices without active displays did not affect the conversation, but those with active displays disrupted the conversation as users lost eye contact.

Our work complements existing work in this domain by further understanding how to design devices so that information delivery can be made subtle without disrupting face-to-face interactions. Evidence from prior work suggests that adaptation and usage of such technology will not be viewed negatively, however, success of these devices depend on how little they disrupt existing social norms and expectations. We focus on studying two parameters that affect disruption these devices may cause: the mode of information delivery, and the amount of information delivered.

UNDERSTANDING USER INFORMATION ABSORPTION

Our goal in this study was to understand design parameters for augmented reality devices that deliver information without disrupting face-to-face conversations. Specifically, we looked to answer the following two research questions:

RQ1. *How does the mode of delivery and amount of information impact how a person processes information internally while conversing with another person?*

RQ2. *How does the mode of delivery and amount of information impact how detectable to their fellow conversants the receivers are as they internalize information?*

To answer these questions we conducted a controlled laboratory study where information was delivered surreptitiously to a person who is engaged in a conversation with another person. For this study, we recruited pairs of participants to have multiple conversations on topics we pre-selected for them. One person in the pair also received a set of words (without knowledge of the other person) via various delivery modes using a custom device, which they had to process during the conversation. We measured how well participants were able to disguise the fact that they were also processing a separate stream of information during the conversation. Lessons from the study should help generate design guidelines for creating technology in the augmented reality domain that can be used to deliver information without impacting users' ongoing interactions noticeably.

Experimental Design

The study was a 3 (Information delivery mode: auditory, visual persistent, visual non-persistent) X 5 (Length of information: 3, 5, 7, 9 or 11 words) within subjects repeated measures study. Each delivery mode was repeated three times, and within each trial there were two sets of information of each of the 5 lengths that we tested. There were four baseline trials: one for the conversation, and one each for processing words using the three information delivery modes (no conversation). To balance for ordering effects, a latin square design was used for the delivery mode, and the length of information factor was assigned randomly within each trial. Baselines were assigned at the beginning and once every three experimental trials.

Users

Users participated in pairs, each instructed to participate roughly equally in a conversation on topics that we preselected. One of the participants was assigned the role of the 'receiver' who would receive information during the conversation, and the other participant was assigned the role of the 'observer' who would try to detect when the receiver was getting information during the ongoing conversation. We did not mirror the condition because we wanted to save both people from having to adapt to different tasks, which could add confusion, and because of time limitations. 24 pairs of participants who did not know each other and who were native English speakers were recruited for the study using a corporate recruiting service, allowing us to access a diverse population. Their ages ranged from 19 to 49 (M=30, S.D= 7.1) and the gender distribution was as follows: FF=8, MF=5, FM=5 and MM=6, where the first letter indicates the gender of the receiver and the second letter indicates the gender of the observer. Occupations included nanny, health service manager, students, security officer, rapid response engineer, writer, physical therapist and so on. Users were compensated with a free software or hardware gratuity. To incentivize appropriate attention on performing the task, an additional gratuity was offered to the 'receiver' and the 'observer' who had the best performance in the study, measured through points that they could gain during the session (described later).

Tasks

For each trial, the primary task was to carry on a conversation on a pre-selected topic for about 5 minutes. Topics included experiential subjects, e.g., favorite childhood memory, high school graduating class, interesting book read recently, role model, and nightmare travel experiences. Participants were asked to maintain a balanced conversation so that each participant talked roughly half of the time. To ensure that the balance was maintained, we flashed a light to alert them if the conversation became one-sided.

Secondary task for the receiver

As the pairs engaged in the conversation, the receiver also performed an information processing task. For this task, the receiver received a list of independent words (i.e., the words were not part of the same phrase), the set size being between 3 and 11. Table 1 shows a subset of the words being presented during a conversation where the topic was 'hobbies'. The observer was aware that the receiver may receive some information during the conversation, but they did not know when the information was being delivered.

We asked receivers to perform two tasks using the word-lists. For the main task, for each word in the list, receivers had to identify whether the word was relevant to the current conversation topic (or not). The purpose of this task was to make sure they were reading and processing the words to be able to respond to them appropriately. They were provided with a clicker with two buttons corresponding to 'relevant' and 'not relevant' responses (see Figure 1). Selecting a response via the clicker automatically advanced them to the next word in the list, until the last word was displayed.

Receivers were instructed to perform this task while still participating in the conversation and to try to disguise that they were multitasking. To ensure that the receiver was motivated to do the secondary task, we told them that they would gain a point for each correct identification of word relevance. They would also receive bonus points if they were able to use one or more words from the list in the conversation. The purpose of this additional task was to assess the viability of this setup as a conversation assistant.

However, to make sure that the receivers did not sacrifice performance on the main task, they were told that they would lose a point every time they were caught being distracted by the observer, so that they would attempt to balance processing the words and participating in the conversation to not appear distracted.

Word sets were delivered to the receiver at most ten times during each trial, roughly with an interval of 30 seconds between word set arrivals. For each trial, words were delivered via one of the three modes: audio, visual non-persistent and visual persistent, defined as follows:

Audio: Presented pre-recorded sets of words sequentially via headphones every 3 seconds, or right after the receiver processed the previous word; **Visual non-persistent:** Replicated audio mode using a visual medium (described in the

Wordlist Size	Words presented
3	Cooking, radio, peaceful
5	Invigorating, quilting, glowing, fine, snorkeling
7	Lonely, kayak, efficient, inferior, glow, Halloween, depression
9	Swim, movies, anger, performance, fresh, serene, pure, fishing, robotics
11	Friends, video-game, enhanced, addicted, ant-farm horse-back-riding, zoo, embarrassed, smooth, serenity, jaded

Table 1: Example of wordlists delivered during a conversation where the topic was 'Hobbies'.

next subsection), i.e., words appeared at the same location on the visual medium one by one and disappeared after 3 seconds, or when the receiver processed it; **Visual persistent**: Presented the entire set of words arranged vertically all at once, via the same visual medium. The receiver processed each word sequentially by hitting the clicker. Visual information was gradually faded in and out of the display to avoid sudden gaze change of the receivers that may reveal their action to the observers.

Secondary task for the observer

The task of the observer was to detect when the receiver was performing secondary tasks during the conversation. The purpose of this task was to assess how receiving information affected the receivers' ability to hold a normal conversation and how detectable this background processing of information is. Observers did not have knowledge about when or via what medium the receiver was receiving the word lists, therefore they were instructed to identify when the receiver appeared distracted to them. The observers also had a clicker with a single button that they used to indicate when the receiver was distracted (see Figure 1). Table 2 summarizes the conditions for the study.

Prototype system for information delivery

The setup of our study required participants' faces to be visible at all times, as one would expect in a natural face-to-face conversation. At the same time, we needed a mechanism for delivering information surreptitiously to the receiver without the knowledge of the observer, allowing delivery using both audio and visual modes. This is particularly challenging for information delivery using the visual channel, as it requires some surface on which to display information. Current implementations of head-up displays do not meet our requirements as they tend to cover the user's eyes and part of the face. Alternatives such as Google glasses that hardly block the user's face are also not suited for our needs: their displays are mounted at the edge of the user's field of view and as a result, the usage of such displays results in a very visible change of gaze.

To solve this problem, we built a custom teleprompter-like device that enables delivery of information to users in a way that is expected to least disrupt their natural conversation. The device is designed with the goal of allowing a

Trials: 16 (3 dual task practice, 1 primary & 3 secondary baseline, 9 dual task)			
Tasks		Factors Manipulated	
Primary	Secondary	Modes	Size
Both: Converse on a given topic	**Receiver:** • Select which words are related • Insert words into conversation (opt) **Observer**: • Detect distractions	**Audio:** Words delivered sequentially via audio **Visual Non-persistent:** Words delivered sequentially via teleprompter **Visual Persistent:** Words delivered in blocks via teleprompter	3,5,7, 9,11

Table 2: Summary of conditions

person to subtly receive both audio and visual signals during the conversation, which does not explicitly interrupt the ongoing conversation. At the same time it allows participants to see each other without any obstruction. This allows the conversation to still seem natural to their collocutor, as the delivery of the signals does not occlude the receiver's face, facial expressions, and eye gaze.

Figure 1 illustrates the setup. The two participants sit on either side of a screen that is visible to the receiver but transparent to the observer. The device allows information delivery through two modes: visual and audio. When in visual mode, words displayed on a concealed LCD screen are projected on a tilted planar glass. This reflects the image towards the receiver, not allowing any of the light to be refracted in the observer's direction and providing a view angle of 25 degrees for a person sitting 1 meter away. The receiver sees the words superimposed on what they see through the glass. In contrast to a regular teleprompter, the screen is set at a depth (w.r.t. the glass) similar to the distance between the observer and the glass. This results in similar focal distances for both words and the observer's face, making them easier for the receiver to read, while focusing on the observer's face. The displayed words are about 0.5 degrees high with a visible distance of 0.5 degrees between them. The observer sees the receiver through the glass, and cannot see the projected image. The audio signals are transmitted to the receiver via headphones, which are worn by the receiver throughout the study.

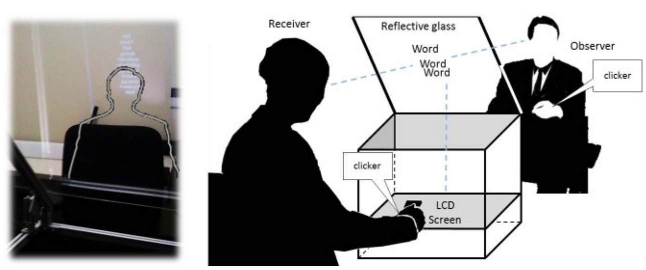

Figure 1. Illustration of the study setup. The receiver and observer sit on either side of our system where information is displayed on a reflecting glass (see actual photo on the left). Both participants have clickers in their hand hidden from the view of the other person. The receiver wears headphones at all times to disguise when information is received via audio.

Both the receiver and the observer hold clickers to register significant events: word relevance for the receiver and distraction detections for the observer. Both clickers have been silenced with a muffling cloth to prevent detections due to the noise of the receiver clicking and other distractions.

METHODOLOGY

On arrival to the lab, participants were assigned roles of 'receiver' and 'observer' according to the experimental design. For trials where both participants were male or both participants were female, roles were assigned randomly. Receivers and observers were then seated at opposite sides of the device (see Figure 1) and taken through an informed consent process. An experimenter read out instructions for the study followed by three practice trials, where participants were exposed to all three delivery modes.

After participants completed the practice trials, they moved on to the experimental trials. The experimental set started with a randomly assigned baseline trial, followed by sets of three experimental and one baseline trial, resulting in a total of 13 trials for the study. For the baseline conversation trial, participants conversed, but no word list was delivered. The observer was unaware of this fact and the purpose of this was to determine the baseline detection of false positives.

For the baseline word list processing trials (one for each delivery mode), 10 word lists (2 for each length) were delivered for each trial, but there was no conversation and the observer was asked to leave the room. The purpose of this trial was to identify the baseline performance for processing words when there was no additional task of conversing or hiding the processing from the other participant.

During tasks, we recorded the conversation, and faces and hand movements of the participants. Time-stamped clicks from the clickers were recorded through our program that was being used to deliver words, so that clicks could be associated with word delivery in post-hoc processing. We also annotated what words from the word list were being used in the conversation in real time, and later validated this with the audio recordings of the conversation.

At the end of each trial both participants filled out separate questionnaires that the experimenters created during the conversation focusing on the content of the conversation. Each person answered questions on what the other person had talked about. The purpose of the questionnaire was to ensure that the participants were attentive to the conversation and not only focusing on the secondary task.

RESULTS

There were a total of 312 scheduled trials including baselines across the 24 pairs. The data for 25 trials were excluded as they either had to be discarded due to equipment malfunction or the trials could not be performed as the participants ran out of time. For some trials not all wordlists were delivered twice and this is accounted for in the analysis using provisions for missing values. A total of 16771 words were delivered, each of which was encoded into a row. The corresponding detections (or lack of thereof) were mapped to each word. An additional 841 rows were generated for false positives in detections using the same format.

We report results according to the research questions, focusing on effects of the delivery mode and bandwidth. For RQ1, we looked at how well the receiver responded to the incoming information while conversing. For RQ2, we looked at how well the observer was able to detect the receiver responding to the information. Unless otherwise stated, for each measure we first report on the differences between dual-task and baseline to establish how dual-tasking affected that metric. We then look at only the experimental trials (where dual tasking occurred) to understand effects of the mode and amount of information delivered on the metric. Table 7 provides a brief summary of findings.

RQ1: Receiver's response to incoming information while conversing

To understand how receiver's ability to respond to incoming information was affected by the mode and amount of information, we looked at three metrics: whether the receiver responded to delivered words or not, the time taken to respond, and whether they were able to insert any of the words that they received into the conversation (the last task was optional for participants).

Whether the receiver responded to a word or not

Comparison between baseline and experimental: A two-way contingency analysis with Response (Responded, Did not respond) and Condition (Baseline, Experimental) found them to be significantly related: Pearson's $\chi^2(2, N=16160)=1149.03$, $p<0.001$, Cramer's V=0.267. In the baseline condition, 87.7% of the words received a response from the receiver, whereas only 61.9% of the words received a response in the experimental condition. This shows that, while conversing receivers chose not to ignore words, perhaps so that the conversation did not get disrupted.

Within experimental trials: For trials where a word list was delivered during a conversation, we conducted two more two-way contingency analyses to determine whether receivers' tendency to respond varied with 1) Delivery mode (audio, visual non persistent, visual persistent) and 2) Length of the word list (3, 5, 7, 9, 11). Delivery mode and Response (Responded, Did not respond) was significantly related: Pearson's $\chi^2(2, N=10816)=306.42$, $p<0.001$, Cramer's V=0.168. Receivers were likely to respond to 70.8% of the words in the Visual persistent condition, compared to the 55.7% in Audio and 53.6% in Visual non-persistent. The relationship between List length and Response approached significance (Pearson's $\chi^2(4, N=10816)=8.896$, $p<0.064$, Cramer's V=0.029. Response percentages were: length 3 (58.8%), length 5 (61.2%), length 7 (60.8%), length 9 (62.8%) and length 11 (63.3%).

Time to respond

For Audio and Visual non-persistent, *time to respond* was measured as the time between when a word was presented to the receiver, and when they hit the clicker to decide whether the word was relevant or not. For Visual persistent, where all words appeared at once, the response time was computed as the time from the last word responded to the response time for the current word. For the first word in the list, the response time was computed as the difference between the list arrival and the response.

Comparison between baseline and experimental: A univariate ANOVA on Response time with Condition (Baseline, Experimental) as a factor showed a significant effect of Condition ($F(1,16159)=150.69$, $p<0.001$): receivers took more time to respond on the baseline (M=1.71s, S.D= 1.68) compared to the experimental trials (M=1.26, S.D.=2.36). This is an interesting observation as it suggests that though receivers responded to a smaller percentage of words (62%) in the experimental trials, they did so quickly, perhaps to get the task out of the way and focus on the conversation.

Within experimental trials: A univariate ANOVA on Response time with Delivery mode and Word list length as factors revealed significant effects of Delivery mode ($F(2,10816)=186.21$, $p<0.001$) on responding to words during conversations. There was also a marginal interaction effect between Delivery mode and Word list length ($F(8,10801)=1.915$, $p<0.053$).

Looking at the interaction effects, there was a significant difference in response time for word lists of various lengths only in the Visual persistent mode. Post hoc Bonferroni tests showed that words belonging to lists of length 3 were responded to faster (M=1.01, S.D. = 2.02) compared to words in lists of length 5 (M=1.29, S.D.=2.28, p<0.419), length 7 (M=1.39, S.D.=2.63, p<0.037), length 9 (M= 1.46, S.D. =2.49, p<0.005) and length 11 (M=1.58, S.D.=2.62, p<0.0001). This result suggests that when multiple words are presented together, more time is required per word to process words in the list. This is further supported in the non-significant difference in response times for words in lists of varying lengths for Audio or Visual non-persistent, where words appeared and were responded to one by one.

For the delivery mode, post-hoc Bonferroni tests showed words appearing in the Audio mode had significantly higher response time (M=1.77s, S.D.=2.7) compared to both Visual persistent (M=1.42, S.D.=0.2.49, p<0.001) and Visual non-persistent (M=0.61, S.D.=1.611, p<0.001). Response time for Visual persistent was also significantly higher than response time for Visual non-persistent (p<0.001). That processing Audio is the most time consuming during conversations is likely due to conflicts in processing two auditory streams simultaneously. Visual non-persistent, which has the quality of Audio in terms of fading out soon after arrival, was processed the quickest, likely because it was easier to process while also conversing and because receivers knew that the word would disappear soon.

	Word was delivered	No word was delivered	Total
Observer identified distraction	1026	841	1867
Observer did not identify distraction	15745	0	15745
Total	16771	841	17612

Table 3: Breakdown of true positives, false positives and missed detection of distraction

Insertion of words into the conversation

Receivers were encouraged to insert words from the word lists into the conversation if possible. The rate of this was low, resulting in a total of 216 words being inserted in the conversation across all users (Audio: M=2.32, S.D.=3.03, Visual persistent: M=3.31, S.D. 3.9, Visual non-persistent: M=2.85, S.D.=3.9). An ANOVA analysis did not reveal any significant effects of the Delivery mode on the insertion, suggesting that inserting the words did not vary based on how the words were delivered to the user. We were unable to associate the length of wordlists to word insertions as often receivers would use words much later in the conversation after its onset. Nonetheless, because word insertions were not the participants' main task, we cannot conclude that participants were unable to insert words. Instead, we observe that it is possible to insert words in a conversation as they are received. We plan to quantify how well users can insert words as their main task in the future.

RQ2: Observer's ability to detect the receiver responding to information during a conversation

To understand how detectable the receiver was while responding to information during a conversation, we looked at the following measures: how often the observer was able to correctly detect the receiver responding to information and how many detections happened based on wordlist size.

Across 24 users and 1516 wordlists, a total of 16771 words were delivered. Observers were able to correctly detect 1026 of these (recall: 6.1%). Observers also detected 841 distractions when there was no word delivered, resulting in a precision of 54.9%. See Table 3 for a breakdown.

Whether observers were able to detect the receiver

Comparison between baseline and experimental: A two-way contingency table showed a significant relationship between Condition (Baseline, Experimental) and Detection: Pearson's $\chi^2(2, N=12169)=366.17$, $p<0.0001$, Cramer's V=0.173. 96.4% of the detections were during the experimental trials and only 3.6% of the detections were during baselines (31 across 24 baseline trials, M=1.29/trial). Out of the detections on the experimental trials, 59% were correct detections (M=3.9/trial), and 41% were false positives (the observer detected the receiver being distracted when no word was presented, M=3.1/trial). The numbers suggest that these detection rates were not due to random clicking as the pattern for the baseline trials (1.29 detections/trial) were much lower than the experimental trials (>3 detections/trial). Also for the experimental trials words were delivered less than half of the time, and random clicking would have yielded a higher percentage of false positives.

	Audio	Visual non persistent	Visual persistent
Correct detection	14.5%	7.4%	6.8%
False positive	3.8%	9.3%	4.9%
Missed detection	81.7%	83.1%	88.3%

Table 4: Breakdown of correct detections, false positives and missed detection across the three delivery modes.

	Receiver speaking	Observer speaking
Correct detections	37.1%	19.4%
False positives	34%	9.5%

Table 5: The effect of speaker on detection correctness.

Within experimental trials: Looking only at trials where a wordlist was delivered, we wanted to see the effects of delivery mode and wordlist length on the correct detections, false positives and missed detections. A two-way contingency table showed a significant relationship between Detection type (correct detection, false positive and missed detection) and Delivery mode: Pearson's $\chi^2(4, N=11729)=247.4$, $p<0.0001$, Cramer's V=0.145. Overall, words delivered via Audio had the highest correct detection rate (14.5% of all words delivered via that mode) compared to 7.4% for Visual non-persistent and 6.8% for Visual persistent. Also, Audio had the least false positives (3.8%) compared to 9.3% for Visual non persistent and 4.9% for Visual persistent. Table 4 summarizes the breakdowns. Groups were not significantly different in terms of detections across word list lengths, so we omit these results.

Number of detections per wordlist

Since we established in the previous subsection that 96.4% detections were during the experimental trials, we do not repeat the baseline comparison here. Results found a significant effect of both Delivery mode ($F(2,1501)=32.2$, $p<0.0001$) and Word list length ($F(4,1501)=26$, $p<0.0001$) on the number of detections per wordlist. For mode, post-hoc Bonferroni tests showed that observers had significantly higher number of detections per wordlist when the wordlist were delivered via Audio (M=1.03, S.D.=1.42), compared to Visual persistent (M=0.54, S.D.=0.88, $p<0.0001$) and Visual non-persistent (M=0.56, S.D.=0.89, $p<0.0001$). For word lists, Bonferroni tests showed that word lists of length 3 had significantly lower detections (M=0.322) compared to length 11 (M=1.124), length 9 (M=0.848), length 7 (M=0.723) and trended lower than length 5 (M=0.538). However, this is not unexpected, as the higher the number of words in a list, the longer the time it is displayed, and thus the higher the probability that there will be more detections. For this, we normalized the detection counts according to the word list length and found that there were no significant differences in normalized detection counts across word lists of different lengths.

As Visual persistent was the only mode where all words in the list were shown at once, we focus only on this list. Though there was still no significant differences in the normalized detection count, the values were trending upwards with the length of the wordlist (M(length 3)=0.069, M(length 5)=0.078, M(length 7)=0.074, M(length 9)=0.068 and M(length 11)=0.086) – suggesting that receivers become more detectable as number of words increase.

Additional analysis: Effects of who is speaking

Although not in the manipulated factors, during the study it appeared that performance was affected by who was speaking at a given moment. To further explore this, we looked at two metrics (only for experimental trials where speech happened during processing of words) – whether a receiver responded to a word or not, and detection by an observer.

For receivers' response to words, a two-way contingency table analysis showed significant relationship between Response and who was speaking: Pearson's $\chi^2(1, N=10712)=98.99$, $p<0.001$, Cramer's V=0.96. Receivers responded to a word 66.2% of the time if the word arrived while the observer was speaking, and 56.8% of the time if they were speaking themselves. This suggests that receivers are more inclined to process information if it arrives when they are not actively engaged in the conversation.

For detections by the observers, looking at all detections, a two-way contingency table analysis suggested a significant relationship between Detections and Speaker (observer, receiver): Pearson's $\chi^2(2, N=1816)=34.7$, $p<0.0001$, Cramer's V=0.138. This did not vary across delivery mode. We further break correct and incorrect detections down in Table 5. These results show that when the receiver is speaking, observers generate more detection events. However, about half of these detections are incorrect detections (i.e., the receiver was not receiving any information), showing that observers were essentially generating detecting events at random. On the other hand, when the observer speaks, the number of overall detections goes down, but the accuracy of detections goes up: observers are roughly twice as likely to be correct in their detections as they are to be incorrect. This is probably due to observers splitting their attention between the task of speaking and the task of detecting; observers mostly detected receivers when receivers gave stronger cues that they were receiving information.

Subjective feedback

At the end of the trials both receivers and observers answered a final questionnaire asking about their experience in the study. Only one receiver was unable to fill out the questionnaire as the server crashed. Receivers were asked to rank-order the delivery modes in terms of how easy it was for them to carry out the word processing task surreptitiously while delivered via that channel. Examining participants' relative rankings of modality shown in Table 6, we observe that Audio was the least preferred mode by more than half of the receivers. However, 34.8% rated Audio to be their top choice. Visual non-persistent was the top choice (43.5% preferring it the most), and it was also a much more popular second choice than Audio was. We believe this is the case because participants perceived it as

Mode	Most preferred	Second choice	Least prefered	Rating average
Audio	34.8%	8.7%	56.5%	1.78
Visual non persistent	43.5%	57.1%	4.8%	2.39
Visual persistent	23.8%	39.1%	39.1%	1.83

Table 6: Modality ranking by receivers.

easy to process due to offering only one word at a time, and in a medium that does not conflict with the conversation in terms of delivery channel. Interestingly, even though quantitative results for Visual Persistent show high performance for this mode, participants were least attracted to it.

Subjects also provided justifications for their choices. The reasons for ranking Audio low included: "*it was difficult to listen to both at once. I preferred one at a time so I didn't have to memorize a couple words while I was trying to talk to someone*", while justifying why Visual non-persistent was preferred over Audio and the Visual persistent; and "*in the visual tests, I was trying to keep eye contact but the words didn't line up with XX's eyes, so I would have to break eye contact... that was the least distracting*, suggesting that the single word lined up with the observer's eyes allowed receivers to maintain eye contact with the observer while processing the word in the background. 12/23 receivers who completed the survey said that the secondary task being presented via Audio was challenging, as it was difficult to process two auditory streams simultaneously. This was also reflected in the quantitative findings, where receivers failed to respond to almost half of the words delivered via Audio. However, preferences for Audio were also voiced. The most common reason is echoed in this comment: "*It was easier to look like I was paying attention when I didn't have to read text. The text lists were most difficult because I felt more bombarded with information …*".

Receivers also revealed strategies they used to disguise their processing an additional task during the conversation. 10/23 receivers said that they tried to maintain eye contact as much as possible. A few others commented that it was difficult for them to hide that they were performing another task and they tried to use filler words such as 'um' and 'ah's to create the impression that they were paying attention.

Observers were asked about what strategies they were using to detect the receiver being distracted. 17 out of the 24 respondents reported that they were looking for eye contact, which matches the strategy that many receivers reported to use to show that they were paying attention. Some observers commented that they looked for unusual breaks in the receiver's conversation. Others commented on leveraging hand movements (as the receiver had to use to the clicker to respond to a word) or overall body language that suggested that the receiver was distracted. As our quantitative results showed observers to be less likely to miss an auditory distraction, it is possible that the difficulty in processing the audio may be resulting in subtle distraction cues via body language that observers were picking up on.

Overall, responses on the questionnaire suggested both receivers and observers had comparable strategies in trying to display a veneer of paying attention and trying to determine when the receiver was not paying attention. Eye contact was the most common technique to show attention or to determine the lack thereof. These findings also suggest that any display providing information without the knowledge of others around the receiver should not require the receiver to look away or lose eye contact, or only display information when this has already happened for some other reason.

Receivers and observers also received questionnaires at the end of each trial, each answering 3 questions on the topic just discussed. There was no significant difference in their scores compared to the baseline when no word list was delivered or in the scores based on the modality of word list.

DISCUSSION, LIMITATIONS AND FUTURE WORK

We now examine the implications of the results we obtained on the design of augmented reality artifacts. Regarding modality, we observed that wordlists delivered to receivers via Audio results in low response rates and high response times from the person receiving, and high number of correct detections from the person observing. Participants performed especially poorly when they were speaking and received words via Audio. This is an interesting result, as one may assume that concealing receipt of information via audio may be easier, especially since no eye movements are necessary as in the visual modes. The cocktail party

Research Question	Measures	Effects of factors tested	Effects of who is speaking	Subjective feedback
RQ1	How receivers responded	Visual persistent yields most responses, longer wordlists are more likely to be responded to.	More inclined to process information if the other person is speaking	Visual persistent is most and Audio is least preferred
	Time to respond	Visual non persistent had shortest response time; words in shorter lists got fastest response		
	Insertion of words	No effect of factors		
RQ2	Can observers detect receivers?	-59% correct detections -Most correct during audio	More correct detections if the other person speaks	Receivers try to maintain eye contact, observers look for lack of contact
	Number of detections per wordlist	Audio is detected the most, shorter wordlists are detected the least		

Table 7: Summary of main findings

effect [4] also suggests that people may be able to process two simultaneous streams of information if relevant to an ongoing conversation. However, an explanation may be found in limitations of cognitive resource sharing - processing two streams coming in via the same channel has been shown to be difficult [31]. We advise against using Audio to convey information during conversations.

The Visual non-persistent mode resulted in low response rates, low response times, and number of correct detections lower than Audio and similar to the Visual persistent mode. We attribute the low response times to users trying to quickly dismiss words being displayed. The lower detection rates suggests that the Visual non-persistent mode demands less from receivers than Audio, resulting in less discernible cues from the receiver that the observer can pick up on.

Finally, the Visual persistent mode resulted in high response rates, medium response times that grow with word list size, and low detection rates that also grow with word list size. We attribute the higher response rates to the "batching" nature of the Visual persistent mode, where users can read and process multiple words at a time.

Although quantitatively users performed better in the Visual persistent mode, users ranked Visual non-persistent mode as the easiest to cope with. One caveat is that the reported discomfort with the Visual persistent mode could have been caused by how we structured the question to participants: unfortunately, we did not ask them about their comfort level with specific word list sizes. It is important to note that even though users perceived the Visual non-persistent mode as easiest to cope with, they were not detected more often in Visual persistent mode, and had higher response levels. This means that users are capable of adapting to the display of multiple words and possibly other delivery modes that require changes of gaze. Additionally, users also frequently reported that they used eye contact to show (in the case of receivers) and detect (observers) whether they were engaged in the conversation. This seems to be in contradiction with the previous observation, but it seems that receivers found strategies to effectively disguise the secondary task, even if they were not comfortable with them. These observations seem to imply that a Visual persistent mode with a low number of words is a good trade-off. As future work, we plan to investigate whether the positioning of these words has an impact on user satisfaction with the system.

Finally, we observed a significant difference in performance of both receivers and observers depending on who was talking when words were delivered. When receivers were speaking, they were less likely to respond. Additionally, in these situations, observers were more likely to detect receivers getting the words, but only because they were more likely to click (true detections were as common as false positives). When observers were speaking, however, receivers were more likely to respond. In these situations, observers detected receivers fewer times, probably because their focus was on the task of speaking. They were however more precise in their detection, likely because it took more significant cues from the receiver to call for their attention. Still, the number of overall correct detections was lower when observers were speaking, so we recommend that words are delivered to receivers when they are in silence.

Asking receivers to optionally insert words in the conversation may have possibly distracted them from the mandatory secondary task, which was indicating whether they thought a word was related to the topic of the conversation. However, we decided to include this task to demonstrate that it is possible for a user to use words delivered during a conversation. Even though that was not the focus of our study, our results show evidence that this is possible.

It is worth noting that our user recruitment may have biased our results, as we only recruited native English speakers. The reason for this was that we wanted to factor out the influence of mother tongue on distractions or loss of detection accuracy. One strong trend we found was that participants assumed eye contact is a strong signal of attention, which is the case in the American culture, but may not generalize to other cultures. However, extending the findings to other populations is beyond the scope of this paper.

The usage of a transparent custom teleprompter met our needs for uninterrupted view of the participants, and the scenario we investigate. It could be used to evaluate other AR scenarios such as those where clerks or sales people provide services via windows (e.g., information that could enable more personalized service to customer projected on the window glass), and video conferencing scenarios.

A central assumption in this work is that distraction during face-to-face conversations is detrimental and thereby should be minimized. Part of this assumption is, as the collocutor is not aware of what their fellow conversant is distracted with, judgment about their lack of attention towards the conversation is possible. Would people be more open to others visibly consuming information if they knew that the information is either important to the conversation or urgent for the person receiving it? Prior work has shown that shared goals can help people better manage interruptions in terms of being more considerate about when to interrupt the other person [9, 24] – here, we can imagine that with the shared goal of a more informed and involved conversation, it will be more socially acceptable to receive information while conversing, as long as the other person is aware of this. Future work will investigate these possibilities.

CONCLUSION

We presented a study that revealed parameters that designers should consider when developing information delivery interfaces for new augmented reality devices. Results show that users can process information while conversing without being detected by their conversation partners. They perform best when presented with small batches of visual information and when they are not speaking. These findings can inform design of devices that deliver just-in-time information to people engaged in other tasks, e.g., face to face conversations.

ACKNOWLEDGMENTS

We thank Tim Sherwood for initial discussions and brainstorming, and Hrvoje Benko and Andy Wilson for their valuable feedback on this work.

REFERENCES

1. Albanesius, C. Google 'project glass' replaces the smartphone with glasses, http://www.pcmag.com/article2/0,2817,2402613,00.asp,2012.
2. Argyle, M. and Cook, M. *Gaze and Mutual Gaze*. Cambridge University Press, London, 1976.
3. Bateman, S., Teevan, J. and White, R.W. The Search Dashboard: Changing How People Search Using a Reflective Interface, *CHI* 2012, 1785-1794.
4. Bronkhorst, A.W. The Cocktail Party Phenomenon: A Review on Speech Intelligibility in Multiple-Talker Conditions. *Acta Acustica 86, 2000*. 117-128.
5. Brumby, D.P., Davies, S., Janssen, C.P. and Grace, J.J. Fast or safe?: how performance objectives determine modality output choices while interacting on the move. *CHI* 2011, 473-482.
6. Brush, A.B., Meyers, B.R., Scott, J. and Venolia, G.D. Exploring Awareness Needs and INformation Display Preferences Between Coworkers *CHI* 2009, 2091-2094.
7. Cakmakci, O. and Rolland, J. Head-worn displays: a review. *Journal of Display Technology*, 2006, *2* (3). 199-216.
8. Campbell, S.W. and Kwak, N. Mobile Communication and Civil Society: Linking Patterns and Places of Use to Engagement with Others in Public. *Human Communication Research*, 2011, *37* (2). 207-222.
9. Dabbish, L. and Kraut, R.E., Controlling interruptions: awareness displays and social motivation for coordination. in *CSCW* 2004, 182-191.
10. Goffman, E. *The presentation of self in everyday life.* Doubleday, Garden City, NY, 1959.
11. Gowans, G., Campbell, J., Alm, N., Dye, R., Astell, A. and Ellis, M. Designing a multimedia conversation aid for reminiscence therapy in dementia care environments. *CHI* 2004, 825-836.
12. Iqbal, S.T., Grudin, J. and Horvitz, E. Peripheral computing during presentations: perspectives on costs and preferences, *CHI* 2011, 891-894.
13. Jansen, B.J. Seeking and implementing automated assitance during the search process. *Information Processing and Management*, 2005, *41*, 909-928.
14. Ko, A.J., DeLine, R. and Venolia, G. Information Needs in Collocated Software Development Teams *International Conference on Software Engineering*, 2007, 344-353.
15. Maglio, P. and Campbell, C.S., Tradeoffs in Displaying Peripheral Information. *CHI* 2000, 241-248.
16. Mann, S. Wearable Computing: A First Step Towards Personal Imaging. *IEEE Computer*, 1997, *30* (2). 25-32.
17. McAtamney, G. and Parker, C. An examination of the effects of a wearable display on informal face-to-face communication. *CHI* 2006, 45-54.
18. 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. *IJHCS*, 2003, *58* (5). 547-582.
19. Milrad, M. and Spikol, D. Anytime, Anywhere Learning Supported by Smart Phones: Experiences and Results from the MUSIS Project. *Educational Technology and Society*, 2007, *10* (4). 62-70.
20. Papagiannakis, G., Singh, G. and N.Magnenat-Thalmann A survey of mobile and wireless technologies for augmented reality systems. *Journal of Visualization and Computer Animation*, 2008 *19*. 3-22.
21. Rash, C.e.a. The Effect of a Monocular Helmet-Mounted Display on Aircrew Health: A Cohort Study of Apache Ah Mk1 Pilots Four-year Review, United States Army Aeromedical Research Laboratory, 2010.
22. Rauterberg, M. Ubiqutious and Mobile Computing - New Directions in User System Interaction. *IEEE Proceedings Symposium on Wireless Multimedia, 1998*.
23. Rhodes, B., Minar, N. and Weaver, J. Wearable computing meets ubiquitous computing: Reaping the best of both worlds *ISWC* 1999, 141-149.
24. Romero,N., Szostek,A., Kaptein,M. and Markoloulos,P. Behaviours and preferences when coordinating mediated interruptions: Social and system influence *ECSCW*, 2007, 351-370.
25. Roussos, G., Marsh, A.J. and Maglavera, S. Enabling Pervasive Computing with Smart Phones. *IEEE Pervasive Computing*, 2005 *4* (2). 20-27.
26. Scherl, C.R. and Haley, J. Computer Monitor Supervision: A clinical note. *The American Journal of Family Therapy*, 2000, *28*. 275-283.
27. Shirey, W.J. The Frustration Phenomenon: Exploring Leader-Follower Relationships in the Information Age. *International Journal of Leadership Studies*, 2008, *3* (2). 223-229.
28. Subramaniam, P. and Woods, B. Towards the therapeutic use of information and communication technology in reminiscence work for people with dementia; a systematic review. *Int. J. Comput. Healthc.*, 2010, *1* (2). 106-125.
29. Sutherland, I.E. A Head-mounted Three Dimensional Display *The Fall Joint Computer Conference, part 1, AFIPS '68(Fall part 1)*, ACM, New York, NY, 1968, 757-764.
30. Tamminen, S., Oulasvirta, A., Toiskallio, K. and Kankainen, A. Understanding mobile contexts. *Personal Ubiquitous Comput.*, 2004, *8* (2). 135-143.
31. Treisman, A.M. Contextual Cues in Selective Listening. *Quarterly Journal of Exp. Psychology*, 1960, *12*. 242-248.
32. vanKrevelen, D. and Poelman, R. A Survey of Augmented Reality Technologies, Applications and Limitations. *The International Journal of Virtual Reality*, 2010, *9* (2). 1-20.
33. White, R.W. and Buscher, G. Characterizing Local Interests and Local Knowledge, *CHI* 2012, 1607-1610.
34. Wiese, J., Biehl, J., Turner, T. and van Melle, W. Beyond 'yesterday's tomorrow': Towards the design of awareness technologies for the contemporary worker, *MobileHCI*, 2011, 455-464.