The Attentive Hearing Aid: Eye Selection of Auditory Sources for Hearing Impaired Users

Jamie Hart, Dumitru Onceanu, Changuk Sohn, Doug Wightman and Roel Vertegaal

Human Media Lab
Queen's University
Kingston, Ontario
Canada K7L 3N6

Abstract. An often-heard complaint about hearing aids is that their amplification of environmental noise makes it difficult for users to focus on one particular speaker. In this paper, we present a new prototype Attentive Hearing Aid (AHA) based on ViewPointer, a wearable calibration-free eye tracker. With AHA, users need only look at the person they are listening to, to amplify that voice in their hearing aid. We present a preliminary evaluation of the use of eye input by hearing impaired users for switching between simultaneous speakers. We compared eye input with manual source selection through pointing and remote control buttons. Results show eye input was 73% faster than selection by pointing and 58% faster than button selection. In terms of recall of the material presented, eye input performed 80% better than traditional hearing aids, 54% better than buttons, and 37% better than pointing. Participants rated eye input as highest in the "easiest", "most natural", and "best overall" categories.

Keywords: Eye Tracking, Attentive User Interface, Assistive Technology, Hearing Impairment, Input Devices, Multi-Modal Input.

1 Introduction

The most common reason cited by hearing impaired individuals for rejecting the use of a hearing aid is intolerance of the large amount of background noise associated with such devices [14]. Traditional hearing aids amplify all sounds in the user's environment, whether the user is interested in them or not [5]. The problem of unwanted background noise has been shown to result in the avoidance of social situations, as well as negative physiological and psychological behavioral changes in users [15]. Over 80% of potential hearing aid wearers opt out of using a hearing aid altogether, reporting this as their chief reason [14]. Our Attentive Hearing Aid project hopes to address this problem by allowing users to target only the voices they wish to listen to, while attenuating background noise.

The technology behind AHA is based on ViewPointer calibration-free eye tracking [28]. It features a small wearable camera pointed at the eyes, which senses when users are looking at one of several infrared tags. These tags are mounted on lapel microphones that are handed out and worn by interlocutors during a conversation. It is our hope that the ability to focus on individual speakers and sound sources allows AHA wearers to enjoy a better quality of life than they would with a traditional hearing aid.

Other directional hearing aid technologies exist, such as the Phonak Directional Hearing Aid [22], that help users direct their hearing towards one particular speaker in multiparty conversations. These technologies typically rely on a directional microphone mounted on the hearing aid, that is pointed toward the sound source via head orientation. While it has been shown to improve intelligibility of speech in simulated room conditions, benefits of directional hearing aids are limited, with one study quoting improvements in speech intelligibility of only about 20% over omni-directional aids [22]. The reason for this is that like omni-directional aids, directional hearing aids are not equipped to cut out extraneous environmental noise sources entirely. Instead, AHA switches microphone sources directly, allowing full control over potential sources in or outside of (visual) attention.

We designed a preliminary study into the performance of AHA as a mechanism for switching, rather than as a specific embodiment of any particular hearing aid technology. We did not include head orientation in our original study for several reasons. Multiple studies have confirmed that eye movements precede head movements when targets are not predictable by time and location [3]. The real world use cases and study addressed by this paper feature target selection that is not predictable by time and location. Head orientation is also known to be inaccurate in target selection tasks [19], users tend not to move their heads when visual targets are within 15 degrees of one another [7], and head orientation relies on neck muscles that are known to be some of the slowest in the human body [16]. The discussion section of this paper describes a subsequent study that further addresses head orientation input. In this study, we focused on comparing eye input with the switching of auditory sources via manual pointing, as well as with the kind of manual switches widely available in hearing aid controls. Performance of these manual input devices is known to be superior to that of head orientation, and we expected eye input to be superior to these manual inputs. The paper is structured as follows. First, we discuss existing literature on hearing impairment and on eye-based selection, after which we describe our prototype AHA. We then discuss our preliminary evaluation. We conclude by discussing possible implications of this technology for both hearing-impaired and normal hearing individuals.

2 Hearing Impairment

A recent survey [14] estimated that there are 31.5 million Americans with some form of hearing impairment, equaling about eight percent of the population. By 2050, it is estimated that about 50 million Americans will suffer from hearing loss [14]. Financially speaking, untreated hearing impairments cost the U.S. economy roughly $56 billion dollars: by way of medical care, lost productivity and special education and training [20].

2.1 Hearing Aids

A hearing aid is defined as “a compact electronic amplifier worn to improve one’s hearing, usually placed in or behind the ear” [12]. Hearing aids work by amplifying (and sometimes altering) sounds in the environment in order to compensate for the malfunctioning anatomy in the ear itself. Although there are several different styles of hearing aids on the market, most devices have the same four basic components [12]:

- A microphone that receives sounds in the environment and converts the sound into an electrical signal;
- An amplifier that makes the signal louder;
- A speaker that outputs the amplified signal into the ear;
- A small battery that powers the electrical parts of the hearing aid.

Research has shown that the use of hearing instruments dramatically improves the quality of life for hearing-impaired individuals. A survey of more than two thousand hearing impaired people found that hearing instrument users were more socially active, more emotionally stable, and both physically and emotionally healthier than non-users with a hearing loss [15]. A particularly disturbing reality is the fact that only one in five people who could benefit from a hearing aid actually wears one [14]. This begs the question: Why do so few hearing-impaired people take advantage of this technology?

2.2 Background Noise

One important reason is the inherent problem of standard hearing aids: they amplify everything in the environment, from useful sounds (voices, televisions, radios) to irrelevant sounds (background chatter, air conditioners). Although it is not entirely clear why, most hearing aids do not appear to allow users to separate the sounds that they want to focus on from unwanted sounds in the environment. Research has shown that the presence of background noise negatively affects people in terms of attention tasks, recognition, reaction time, and verbal memory [26], as well as blood pressure, heart rate, skin temperature, and hormone release [17]. Currently, there are three main ways of reducing background noise in commercial hearing aids:

- **Personal FM Systems.** Personal FM systems consist of a portable microphone that is placed near the person who is speaking, and an FM receiver worn by the hearing-impaired individual [10]. The microphone broadcasts a signal on a special frequency, which is picked up by the receiver. The receiver can either connect to the hearing aid via an induction loop, or can be used with a headset. These systems are very useful in settings where there is just one sound source; for example classrooms, churches, and cinemas. However, there may be issues with interference when multiple FM systems are used in the same location, and it can be difficult to select between multiple sources.
- **Directional Hearing Aids.** Directional hearing aids function by comparing the input from microphones at two (or more) different locations on the hearing aid. By summing the sound signals received from the multiple microphones, sounds in front of the user are emphasized and sounds from the sides or rear of the user are reduced. Directional hearing aids work on the assumption that most desirable sounds will be in front of the user. Research has shown that speech understanding in noisy environments can be improved in this way [30]. However, when noise and signal are diffuse, these hearing aids perform no better than conventional hearing aids [21]. Researchers recently unveiled a pair of “hearing-glasses” [31] that work similarly, with a total of eight microphones embedded in the arms.
- **Digital Noise Reduction Hearing Aids.** Digital noise reduction hearing aids take advantage of the frequency of speech, rather than its direction. Human speech has a frequency range of approximately 200 to 8000 Hz, and the range for common sounds is even narrower. Hearing aids equipped with digital noise reduction work by reducing sounds that fall outside of the frequency range of speech. There are two cases when these systems break down: when the noise falls in the same frequency range as speech and when the noise itself is unwanted speech [20]. Since background speech is the most difficult type of noise for humans to filter out [18], this is a very serious issue.

3 Eye-Based Selection

A large body of research exists on the use of eye input for selection tasks, both on-screen as well as in the real world. As detailed in [6], there are many arguments for the use of eye gaze for focus selection in hearing aids:

- The use of eye movements requires very little conscious or physical effort [13].
- Eye gaze is used in human-human communication to indicate whom the next speaker should be [34], and correlates very well with whom a person is listening to [34]. Hearing aid users would likely already be looking at a speaker, for example, to gauge responses or perform lip reading.
- Eye input prevents overloading of the hands because the eyes form a parallel input channel. It does not require hearing aid users to hold a pointing device.

- Eye movements precede head movements when targets are not predictable by time and location [3].
- Eye movements are much faster than either hand or head movements [7, 8].

Eye input also has its issues. Eye trackers are still expensive, requiring calibrations and bulky head gear when used in mobile scenarios [7]. However, new portable calibration-free technologies such as ViewPointer [28] have become available that address many, if not all of these problems. The Midas Touch problem is often cited in literature [24]. In auditory focus selection, it is easily avoided by not allowing binary selection: Using subtle amplification of selected sources, attenuating other sources of audio, eye input mimics the attentive mechanisms of the brain. Eye contact is known to trigger the very attentional processes that allow focusing on conversations [34], making eye input a natural technique.

3.1 Alternative Approaches

There have been many systems built upon the premise of using head orientation as a source of audio selection; we highlight only a small selection of systems in this review. Eye-R [23] is a glasses-mounted device that stores and communicates information based on both eye movement and infrared (IR) LEDs positioned in the environment. Users wear an IR receiver, and a transmitter that broadcasts a unique IR code. The receiver allows the system to determine when the wearer's head is oriented towards another user or device in the environment.

The Visual Resonator [37] is a recent project designed to be a realization of the so-called "Cocktail Party Phenomenon" first described by Cherry [5], and defined as "the ability to focus one's listening attention on a single talker among a cacophony of conversations and background noise" [1]. Similar to [2], Visual Resonator is an auditory interface that allows the user to hear sound only from the direction that she is facing. The device consists of a pair of headphones with a microphone and an infrared transmitter and receiver mounted on top. Visual Resonator is direction-sensitive because both the transmitter and receiver are always oriented in the direction that the user is facing. Incoming signals are received by the IR receiver, and then sent to the headphones where they are translated into sound. Outgoing speech sound is recorded by the microphone, translated into an infrared signal, and then beamed into the environment.

Most of these systems do not track where the user is actually looking. Therefore, if the user's head was oriented towards Person A, but he was actually listening to Person B; the system would incorrectly infer that the user was listening to Person A. This problem can be addressed by selecting targets using eye rather than head movement.

3.2 Performance of Eye Input

There have been many studies on the use of eye tracking as an input device for targeting in human-computer interaction. Ware and Mikaelian [35] compared three eye pointing styles for selecting targets on a CRT: (1) dwell time click, where the target was selected if the observers' gaze fixated on it for more than .4 s; (2) screen button, where the observer had to fixate on a button on the screen after looking at the target; and (3) hardware button, where the observers pushed a keyboard button while fixating on the target. Results showed click times compared favorably to those of the mouse, with an intercept approximately twice as small. Wang et al. [36] discussed an evaluation of eye-based selection of Chinese characters for text entry. In their task, users chose one of 8 on-screen Chinese characters by looking at the character while pressing the space bar. Results showed eye-based selection was not significantly faster than traditional key-based selection. They attributed this to the fact that the overall time required to complete their task was dominated by decision time, rather than movement time.

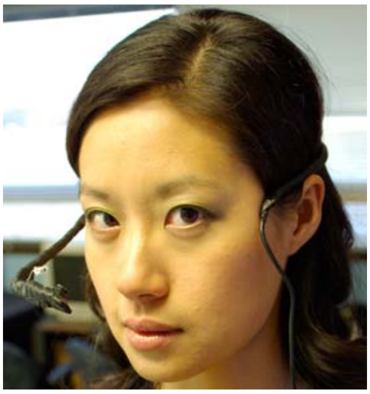

Fig. 1. The camera headpiece.

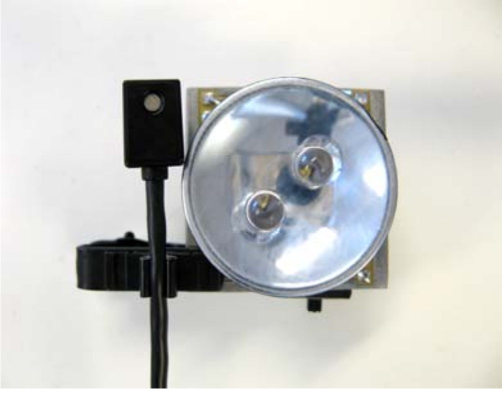

Fig. 2. Lapel mike with wireless infrared tag.

Zhai et al. [38] evaluated the use of eye input in gaze assisted manual pointing. In their MAGIC pointing technique, an isometric joystick was used to select targets on a screen. However, to speed up isometric pointing, they positioned the cursor at a location close to the current eye fixation point whenever the user initiated movement with the joystick. MAGIC pointing only marginally improved movement time in a Fitt's Law task. Sibert and Jacob [24] evaluated the use of a mouse and eye tracker with dwell-time activated click in a pointing task that involved selecting one of 16 circles on a screen. They found that trial completion time with the eye tracker was almost half that of the mouse. EyeWindows [8] compared the use of hotkeys for selection of windows with that of the mouse and two eye input techniques. Results showed that on average, eye input was about twice as fast as manual techniques when hands were overloaded with a typing task. LookPoint [6] is a system that allows for hands-free switching of input devices between multiple screens or computers. A multi-screen typing task was used to evaluate a basic version of the system, comparing eye input with multiple keyboards, hotkeys, and mouse. Results showed that eye input was 111% faster than the mouse, 75% faster than function keys, and 37% faster than multiple keyboards. User satisfaction surveys generally show that participants prefer using the eye input techniques over manual conditions. However, eye input is often inaccurate, with a proneness to wrongful selection.

4 The Attentive Hearing Aid

Our prototype Attentive Hearing Aid (AHA) consists of a wearable infrared camera, which is pointed at the user's eye and connected to a wearable Sony U70 computer (see Figure 1). The user also wears one or two hearing aid ear piece(s), while interlocators wear a lapel microphone that is augmented with an infrared tag (see Figure 2). These lapel mikes broadcast to a portable lapel receiver/mixer and/or induction loop system that connects to the ear piece(s) of the AHA wearer. The wearable camera is based on ViewPointer, a calibration-free eye sensor [27, 28]. The U70 processes images from this sensor and determines whether the corneal reflection of an infrared light source, in this case the tag on the lapel mike, is central to the pupil. This works as long as the camera is within 45 degrees of the visual axis of the eye. For a detailed technical discussion of ViewPointer, please refer to [27, 28]. The battery-powered IR tags on the lapel microphone (see Figure 2) consist of a bank of LEDs triggered by a microcontroller programmed with a binary code that allows tags to be identified through computer vision. When an AHA wearer is looking at another person wearing a tagged lapel microphone, AHA can thus identify the sound source and select it for amplification. Note that microphones need not be lapel microphones, but can be Bluetooth headsets commonly worn by users.

4.1 Video Conferencing Applications

If one does not accept that the above hardware is sufficiently practical for use in day-to-day conversations, there are many other cocktail-party situations, such as multi-person video conferencing, where participants are expected to have a microphone, and are represented within the confines of a screen. Due to the novelty of the equipment, and the large potential for measurement artifacts in a real-world setting, we used such a video conferencing setting during our experimentation.

5 Preliminary User Study

Initial evaluation consisted of hearing-impaired participants selecting targets on a videoconferencing screen in four conditions: ViewPointer input, pointing with a Nintendo Wii remote, button selection with a Nintendo Wii remote, and a control condition that simulated the use of an omni-directional hearing aid. Please note that this evaluation was preliminary, and aimed at determining the fastest and least disruptive technique for switching between sound sources. The participants' task was to follow a story told by three recorded actors on the screen, simulating turn taking in a videoconference. Every ten seconds, a switch occurred, and participants were required to select a new speaker as a source. This was done to mimic natural turn taking behavior in conversations in a controlled setting. We measured the time taken for the participant to select the new target (switch time) as well as participants' recall of the material presented in the story. We chose recall because it is indicative not just of the intelligibility of speech in various conditions, but the actual ability to comprehend that speech. For example, we expected this measure, but not intelligibility, to be sensitive to the mental load caused by the switching mechanism.

5.1 Recall Passages

We created nine passages on topics such as environmental issues, strange animals, and famous women in history. The passages were carefully structured to contain seven target items for recall testing at the end of each trial. Recall was thus measured on a scale from 1 to 7 after each trial.

We used seven of the nine passages as "foreground" stories in our actual experimentation, with the other two stories being used to simulate background conversation and noise. For each of our seven foreground passages, we created a seven-item multiple-choice questionnaire in order to test the recall of the material presented in that particular trial.

5.2 Task

The participant's task was to simply listen to a specific story told by the actors. The actor currently telling the target story was indicated with a red dot below his window. Each trial began with the audio of the target actor already amplified (i.e., a volume of 100%). The other two actors also spoke simultaneously, but at a reduced volume (of 10%) to simulate background chatter. After 10 seconds, the target randomly switched and the red dot moved to the next actor telling the story. The participant then had to select this new actor. This switching procedure repeated every 10 seconds for the entire duration of the trial. After each trial, the participant was immediately presented with a multiple-choice recall test. Seven recall tests were graded afterwards, and the results used as the measure for our recall variable.

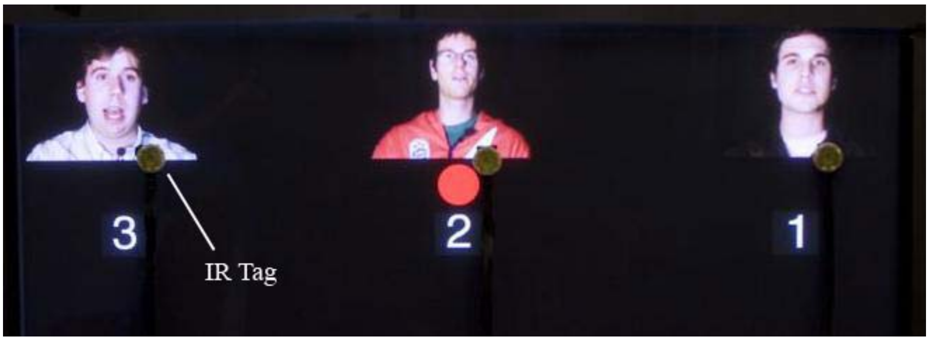

Fig. 3. Training screen in the Buttons condition. The numbers were displayed during training so the participant could memorize the actor-number mapping. IR tags for the eyes condition are also visible as yellow circles right below the actors.

5.3 Conditions: Switching Techniques

We compared eye input with two manual selection techniques, and a control condition consisting of an omni-directional hearing aid. For each participant, we ran one trial of the Control condition; and two trials of each of three experimental conditions. Each participant completed the trials in random order.

- **Control.** In this condition, participants were unable to select which actor they wanted amplified. This essentially represented performance of the most common of hearing aids, the omni-directional hearing aid. For consistency, the red dot still moved from window to window, but the volume of all three actors remained at 100% for the entire duration of the trial.
- **Pointing.** In this condition, participants used a Nintendo Wii remote to point at the target actor.
- **Buttons.** Here, participants used the buttons labeled 1, 2, and 3 on the Wiimote to select the target actor. In a real-world environment, interlocutors would not necessarily stay in one place, thus limiting the usefulness of a location-based button mapping. Participants were asked to select the person instead. Participants were required to first memorize which button (1, 2, or 3) corresponded to which actor before each trial began. We trained participants on the randomized actor-number mapping for a full practice trial before each trial began. During the practice trial, the numbers were displayed on the screen below the actors' window (see Figure 3), and the participants were encouraged to practice switching in order to memorize the mapping. No numbers appeared during the trial, so that participants had to rely on memory. In order to negate learning effects between trials, we used two different mappings in both Button trials.
- **Eye Input.** In this condition, the user simply had to look at the actor whose audio they wanted amplified. The IR tags mounted on the screen registered looks, and the audio of the corresponding actor was amplified within 0.5 seconds.

The participants knew they had selected the correct target actor when the narrative continued correctly across voices. We avoided visual selection feedback as participants would not receive such confirmation in real life scenarios either. Instead, to avoid confusion caused by subjects not recognizing the voice of an actor, in all conditions, audible switches would only take place after the correct target was selected. Thus, participants were allowed to make erroneous selections, but erroneous selection would not allow them to follow the focus conversation.

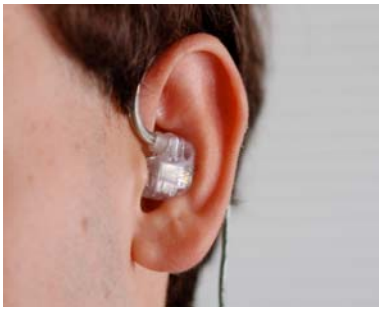

Fig. 4. Shure earpiece with foam insert.

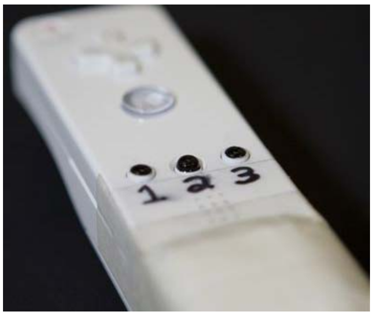

Fig. 5. Wii remote with modification and labeling.

5.4 Participants and Experimental Design

We initially recruited 14 hearing-impaired individuals to participate in our preliminary evaluation. One volunteer had had three major eye surgeries, resulting in a large amount of scar tissue on his eyeball which caused inconsistent corneal tag reflections. Another was suffering from severe allergies and as a result, his eyes were watering profusely, again interfering with the corneal reflection. The third volunteer was hearing-impaired due to a deformity in her outer ear canal, so we were unable to successfully insert earphones. Our final participant group consisted of three males and eight females, ranging from age 13 to age 69 (average age of 48.5 years) (Please note that our subsequent significant findings indicated this sample size was in fact sufficiently large for an initial evaluation). We recruited participants with a wide range of hearing impairments, ranging from people with mild impairment who chose not to wear hearing aids on a daily basis, to people with profound hearing loss who cannot function without their hearing aids. We recruited participants who did not wear glasses, as the current ViewPointer form factor does not accommodate glasses. Each session took approximately one hour, and participants received $10 in compensation. A within-subjects design was employed, with the order of presentation randomized between participants.

5.5 Procedure

At the beginning of each session, the participants read and signed a Consent Form and filled out a questionnaire with details about their hearing impairment. Afterwards, the experimenter explained the procedure, and fitted the participant with sound-isolating earphones. Since hearing impairment varies immensely between individuals, participants first performed a short audio calibration task that we built into our software. We customized the audio by adjusting the overall volume, the balance between the left and right ear, and the relative volume of six different frequency bands. We also took participants' knowledge of their hearing loss into account during this calibration session. Once the briefing and calibration were completed, the experiment began. Before each of the seven randomly selected trials, the participant was allowed to practice with that selection technique until they were comfortable. After the experiment, we presented participants with a five-item questionnaire, asking their opinion on which condition they thought was best for Recall, fastest for Switching, easiest, most natural, and best overall.

5.6 Video and Audio Design

We recorded digital footage of three male actors each reading the nine passages, for a total of 27 movie files. We used movies rather than real conversations because it allowed for greater experimental control, particularly with regards to the turn switching process, that had to occur on a timed basis. We used a timed slideshow to present short phrases at the very top of a computer screen, placing the video camera directly above the screen. Thus actors appeared to maintain eye contact with participants for the entire duration of the recording. It also ensured that the timing and rhythm of the stories was the same between actors. When switching passages every ten seconds the story would continue seamlessly from one actor to the next. Audio was recorded, in stereo sound, with a lapel microphone clipped to the actor's collar. Before trials commenced, participants removed their hearing aids and were outfitted with Shure E5c sound-isolating earphones (see Figure 4) that eliminated all environmental noise. Trials were further conducted in a sound-proof usability lab.

5.7 Screen and Tag Design

Figure 3 shows the screen setup: a 52" plasma screen displaying three videos of actors telling stories. Below each of the three windows was a space where the red dot could appear, which helped participants identify the target window.

We affixed a total of five infrared tags to the screen. Two tags with four infrared LEDs each were centered about 3 cm apart on the upper frame of the screen. These tags emulated the sensor bar in the Nintendo Wii system, emitting solid infrared light to provide triangulation data for the Wii remote pointer. Three ViewPointer tags were mounted on the screen itself (see Figure 3), just below the actors' faces, and were activated in the eye input condition. These tags measured 3 cm in diameter and each consisted of six infrared LEDs. To not have to rely on battery power during experimentation, these tags were connected to a MacBook Pro and directly operated via a Phidget interface board [9]. The ViewPointer camera operated at 30 fps and the tags at 15 Hz. We used the following three eight-bit codes for our tags: 11111110, 11110000, and 10101010. Because we required three tags for this study, the user needed to look at each tag for approximately 0.5 seconds (or 15 frames) before the software recognized and identified the tag. This is well within the average human fixation, between 100 ms and 1 second [32].

During the evaluation, the participant sat in a straight-backed stationary chair that was placed 1.65 m from the screen. We placed a strip of black cardboard over the toolbar at the top of the screen so participants were not distracted.

5.8 Wii Remote

We adapted a Nintendo Wii remote ("Wiimote", see Figure 5) for the two manual input conditions of our evaluation. While we obviously do not expect hearing aid wearers to carry a Wiimote, it is representative for the kind of manual control device that could be considered optimal for this kind of real life task, as it is wireless, operates in 3-space, and relies on orientation only. In real-world conditions, we have observed hearing aid users using a similar small remote control to adjust their settings. Every time after using the remote, they would place it back in their pocket, or on their lap where it could easily be reached. Because it can be tiring to hold up a manual remote control, we allowed participants to relax their arm and place it on their lap in between selections, whenever they got tired of holding up the Wiimote. For the pointing condition, we switched the audio when the participant pointed at an activation box overlaid on the actor's window. For the button condition, we relabeled the –, HOME, and + buttons on the Wiimote to 1, 2, 3 and switched the audio when the participant pressed the correct button. Liquid plastic was applied to the buttons so they would be the same height, and have the same "feel". Figure 5 illustrates the relabeled buttons, as well as the physical modifications made to the Wiimote.

5.9 Software

Max/MSP/Jitter [4] is a graphical programming environment designed for use with multimedia. The program ran in one of three different "modes", with the mode randomly selected for every switch:

1. Same actor/different position – The same actor continues telling the story, but in a different position on the screen. In this case the red dot moves to the new position.
2. Different actor/same window – A different actor continues the story, but in the same window. In this case, the red dot stays in the same position.
3. Different actor/different window – Both the actor and the window position change. As in the first mode, the red dot moves to the new position in this case.

We did not allow for the same actor/same window mode because we needed to force participants to switch every 10 seconds.

5.10 Hardware

Our evaluation required three computers. A MacBook Pro was connected to the USB camera and ran computer vision software to analyze ViewPointer video. The MacBook Pro also ran the Phidget interface used to control and provide power for the infrared tags. The encoded tag number was sent to another PowerBook via the network connection. This PowerBook ran a Max/MSP/Jitter patch with an embedded Java program. This patch then sent out a single integer indicating the focus tag. An iMac ran the Max/MSP/Jitter software that controlled the audio and video switching based on which actor was identified as being in focus. This allowed the iMac to make full use of its processing resources to ensure that the three videos ran seamlessly, and in lip sync.

5.11 Hypotheses

To summarize, our independent variable was switching technique; either (1) None (Control), (2) Buttons, (3) Pointing, or (4) Eye Input. Our two dependent variables were recall (on a multiple choice test with seven questions), and switch time in milliseconds. We hypothesized that switch time would be best with eye input as pointing requires arm movement, and button presses imply a Hick's law selection. As a consequence, we predicted that participants would have the best recall with eye input, and the worst in the Control condition.

5.12 Data Analysis

For each participant, we had a total of six trials of data (two trials each of Buttons, Pointing, and Eye Input). As in [35], we defined switch time as the time between the instant the red dot flashed or appeared in a new location to the instant the user selected the correct actor. The switch time variable obviously could not be applied in the Control condition. For recall, we used the results of the multiple-choice test from each of the seven trials (one trial in Control condition, and two trials each of Buttons, Pointing, and Eye Input). The results indicated the number of correct answers out of a possible seven. Analyses of variance (ANOVAs) with the factor of selection method were performed separately on the switch time and recall variables.

This was followed by post-hoc pairwise comparisons between each condition, using Bonferroni correction to account for multiple comparisons. Questionnaire data was non-parametric and analyzed using Kruskal-Wallis tests. Significance level was assumed at $p < .05$ for all statistical analyses.

Table 1. Mean Switch Time and Recall per condition.

	Control	**Buttons**	**Pointing**	**Eyes**
Mean Switch Time (ms) (s.e.)	n/a	2211.8 (151.6)	2424.6 (166.4)	1404.3 (113.7)
Mean Recall (s.e.)	.82 (.30)	1.91 (.28)	2.60 (.30)	4.14 (.46)

Table 2. User Experience results per condition.

	Control	Buttons	Pointing	Eyes
Perceived Recall	0	5	1	5
Perceived Switch Time	0	4	1	6
Easiest	0	3	0	8
Most Natural	0	1	0	10
Overall	0	3	0	8

6 Results

Results show that Eye Input had a faster switch time than both manual techniques (see Table 1) (F2, 30 = 13.14, $p < .001$). Post-hoc pairwise comparisons with Bonferroni correction showed Eye Input was 73% faster than Pointing ($p < .001$) and 58% faster than Buttons ($p < .05$).

In terms of recall, Eye Input was 80% better than Control, 54% better than Buttons, and 37% better than Pointing (see Table 1) (F3, 40 = 16.33, $p < .001$). Post-hoc comparisons with Bonferroni correction revealed that Eye Input had significantly better recall than Pointing ($p < .05$), Buttons ($p < .001$), and Control ($p < .001$). In addition, recall in the Pointing condition was significantly better than in the Control condition ($p < .01$).

In terms of errors, we found that Eye Input had an average of 6.7 errors per condition, compared to 7 in the Buttons condition and 0 in the Pointing condition. These errors are mostly attributed to participants continuing to hold down a button at inappropriate times in the Button condition, or wrongful detection of a tag activation in the Eye Input condition.

6.1 User Experience

Table 2 shows the results of the five-item User Experience questionnaire across conditions. In terms of subjective ratings, results showed that the AHA was the easiest, the most natural, and the best overall. Kruskal-Wallis tests suggested differences for all five items: perceived recall ($p < .01$), perceived switch time ($p < .01$), easiest ($p < .001$), most natural ($p < .001$), and best overall ($p < .001$).

7 Discussion

Overall, the results obtained for both switch time and recall were in line with expectations. This section presents a discussion of what we believe are the two key explanations for these results: movement time and mental load.

7.1 Movement time as a Limiting Factor

For switch time, Eye Input was 73% faster than Pointing, and 58% faster than Buttons. The chief reason for this is that there is very little movement, and mostly open-loop control involved in eye input. Selecting the actor required little to no cognitive processing, and the only movement required was a saccade. Similarly, in the Buttons condition, the only motion required was a thumb press to activate the correct button. However, it required a mental mapping and a Hick's style 3-way decision, which we will discuss in a subsequent section.

Pointing first typically involves an eye fixation, and a deliberate closed-loop coordination of the wrist and arm movement. We believe this resulted in a longer switch time. Both pointing and button conditions may have been affected by the need to lift up the arm in cases where participants were tired, and rested their arm prior to engaging in selections. However, we believe such bias is, in fact, reflective of real-world limitations of the device.

7.2 Mental Load as a Limiting Factor

In terms of our second dependent variable our results were again in line with our hypothesis; with Eye Input performing 80% better than Control, 54% better than Buttons, and 37% better than Pointing. We believe these results were in part due to switch time as well: the higher the switch time, the more of the story was missed. However, recall was 26% better in the Pointing condition than in the Buttons condition. We believe this difference was due to a higher mental load in the Buttons condition. In that condition, participants could not rely on deixis, or a spatial mapping (i.e., point wherever the dot is) to select the correct actor.[1] Every ten seconds, a Hick's law decision was required: a selection of one of three possible buttons, and a mapping needed to be made between the actor identity and the number to press. We argue this limitation is inherent with any non-spatial device in mobile scenarios, and it presents a distinct limitation for the use of remote control buttons to select persons in real world environments. Participants confirmed that mental load was the lowest in the Eye Input condition.

7.3 Difference between Perceived and Actual Recall

An interesting observation was a large difference between participants' self-reports and their actual data for the recall variable. Participants seemed to overestimate their recall ability in the Buttons condition; and conversely, underestimate their recall in the Pointing condition. One explanation could be that the mental load of remembering the actor-number mapping operates at a subconscious level, and participants therefore were unable to keep track of the extent to which memorization had affected their recall performance. Tognazzini presented a similar argument on perception vs. reality in [29].

7.4 Comparison with Head Orientation Input

Subsequent to the above study, and in response to reviewer feedback, we studied the movement time of selection in the same task using head orientation input. We used a high-resolution webcam tracking a fiducial marker affixed to the participants' heads, tracking their orientation towards the on-screen target. Although due to the post-hoc nature we cannot directly compare data between experiments, trends do indeed suggest that selection through head orientation is slower. Participant recall was about one full point lower than with eye input, which is what we expected, and which is consistent with findings in [3].

7.5 Pros and Cons of the Current Design

We would like to note that this represented a preliminary study aimed at evaluating the potential usefulness of eye input in such scenarios. Further studies are required in the field, in actual face-to-face conditions, and comparing other hearing aid technologies.

One of the main issues with the current AHA design, as well as other directional hearing aids is that the user must always be oriented towards the sound source that they wish to listen to. This means that it would be impossible to listen to the radio next to you, or hear your spouse when she is behind you. The big advantage of AHA over directional hearing aids is that AHA actually selects a sound source, obtaining a signal directly from a lapel microphone, thus eliminating background noise. The use of eye movement rather than head movement also follows more closely what actually occurs in conversations [33], an allocation of brain

[1] Note that this is a fundamental and inherent limitation that was specifically *not* designed to favour pointing or eye input conditions.

resources based on a tuning of visual attention. Our results show it to be much faster than hand movements, which in turn are known to be much faster than neck muscles.

In our evaluation, we chose a controlled environment, which effectively eliminated the Midas Touch Effect [13]. If the Attentive Hearing Aid were to be deployed in the real world, it would always amplify the audio of whoever the user would look at. However, users would also be familiar with the voices of their interlocutors, making wrongful selection easy to detect, and repair is as fast as a fixation on the correct speaker. Hypothesizing an error that occurs in every switch, adding another 500 ms to the mean measure would still find eye input significantly faster than either manual technique.

8 Conclusions

Research shows that the number one improvement sought by hearing aid users is better understanding of speech in noisy conditions. Most hearing aids, including directional hearing aids, have a relatively poor signal-to-noise ratio because they are unable to sufficiently differentiate desired sounds from unwanted noise. We presented the Attentive Hearing Aid, a system that uses eye input from a ViewPointer system to amplify tagged sound sources in the user’s environment. We conducted a preliminary evaluation where hearing-impaired participants were asked to follow a story presented on screen by three actors. Participants selected the target actor every ten seconds in four different conditions: pressing a button, pointing with a remote control, using their eyes, and a Control condition in which actors were speaking simultaneously without filtering. Results suggest that selection with eye input was 73% faster than pointing, and 58% faster than buttons. In terms of recall of presented material, eye input was 80% better than control (no selection/omnidirectional hearing aid), 54% better than Buttons, and 37% better than Pointing. User experience reports were also very positive, with eye input receiving the highest rating in all categories. With proper miniaturization and optimization of components, we believe our results support the tremendous potential for AHA technology to improve the quality of life of users with hearing disabilities in future hearing aids.

9 References

1. Arons, B. Review of the Cocktail Party Effect. Journal of the American Voice I/O Society 12, 1992, 35-50.
2. Basu S. and Pentland A. Smart Headphones. In Ext. Abstracts of CHI 2001. Seattle, 2001, 267-268.
3. Boyer, D.J. The Relationship Among Eye Movements, Head Movement, and Manual Responses in a Simulated Air Traffic Control Task. Technical Report, US Department of Transportation, FAA, 1995.
4. Cycling '74. Max/MSP/Jitter. www.cycling74.com.
5. Cherry, E.C. Some Experiments on the Recognition of Speech, with One and with Two Ears. Journal of Acoustic Society of America 25, 1953, 975-979.
6. Dickie, C., Hart, J., Vertegaal, R., and Eiser, A. LookPoint: An Evaluation of Eye Input for Hands-Free Switching of Input Devices between Multiple Computers. In Proc. of OzCHI'06. ACM Press, 2006.
7. Duchowski, A. Eye Tracking Methodology: Theory & Practice. London, UK: Springer Verlag, 2003.
8. Fono, D. and Vertegaal, R. EyeWindows: Evaluation of Eye-Controlled Zooming Windows for Focus Selection. In Proc. of CHI’05, ACM Press, 2005, 151-160.
9. Greenberg, S. and Fitchett, C. Phidgets: Easy Development of Physical Interfaces through Physical Widgets. In Proc. of UIST'01, ACM Press, 2001.
10. Hawkins, D. B. Comparisons of Speech Recognition in Noise by Mildly-to-Moderately Hearing-Impaired Children Using Hearing Aids and FM Systems. Journal of Speech and Hearing Disorders 49, 1984, 409-418.
11. Hearing Aid. In dictionary.reference.com.
12. Hearing Loss Education Center. www.hearinglosseducation.com.
13. Jacob, R.J.K. The Use of Eye Movements in Human-Computer Interaction Techniques: What You Look At Is What You Get. ACM Transactions on Information Systems 9 (3), 1991, 152-169.
14. Kochkin, S. Hearing Loss Population Tops 31 Million People. Hearing Review 12 (7), 2005.

15. Kochkin, S. and Rogin, C. Quantifying the Obvious: The Impact of Hearing Aids on Quality of Life. The Hearing Review 7 (1), 2000, 8-34.
16. Langolf, G. D., Chaffin, D. B., & Foulke, J. A. An investigation of Fitts' law using a wide range of movement amplitudes. Journal of Motor Behavior, 8, 1976, 113-128.
17. Maschke, C., Rupp, T., and Hecht, K. The Influence of Stressors on Biochemical Reactions - A Review of Present Scientific Findings with Noise. International Journal of Hygiene and Environmental Health 203, 2000, 45-53.
18. Morrison, H.B. and Casali, J.G. Intelligibility of Synthesized Voice Messages in Commercial Truck Cab Noise for Normal-Hearing and Hearing-Impaired Listeners. International Journal of Speech Technology 2, 1997, 33-44.
19. Nickel, K. Stiefelhagen, R. Pointing Gesture Recognition on 3d-Tracking of Face, Hands and Head Orientation. In Proc. ICMI (2003), 380-400.
20. O'Neill, G., Summer, L., and Shirey, L. Hearing Loss: A Growing Problem that Affects Quality of Life. Challenges for the 21st Century: Chronic and Disabling Conditions 2, 1999, 1-6.
21. Ricketts, T.A. and Mueller, H.G. Making Sense of Directional Microphone Hearing Aids. American Journal of Audiology 8 (2), 1999, 117- 128.
22. Ricketts, T.A. and Dhar, S. Comparison of Performance across Three Directional Hearing Aids. American Journal of Audiology 8 (10), 1999, 180- 189.
23. Selker, T., Lockerd, A., and Martinez, J. Eye-R, A Glasses-Mounted Eye Motion Detection Interface. In Ext. Abstracts of CHI'01, ACM Press, 2001, 179-180.
24. Sibert, L.E. and Jacob, R.J.K. Evaluation of Eye Gaze Interaction. In Proc. of CHI’00, ACM, 2000.
25. Smith, A.P. Noise, Performance Efficiency and Safety. International Archives of Occupational and Environmental Health 62, 1990, 1-5.
26. Smith, A.P. Noise and Aspects of Attention. British Journal of Psychology 82, 1991, 313-324.
27. Smith, J.D. ViewPointer: Lightweight Calibration-Free Eye Tracking for Ubiquitous Handsfree Deixis. Master's Thesis, Queen's University, 2005.
28. Smith, J.D., Vertegaal, R., and Sohn, C. ViewPointer: Lightweight Calibration-Free Eye Tracking for Ubiquitous Handsfree Deixis. In Proc. of UIST'05, ACM, 2005, 53-61.
29. Tognazzini, B., Tog on Interface. Addison-Wesley Longman Publishing Co., Boston, MA. 1992.
30. Valente, M., Fabry, D.A., and Potts, L.G. Recognition of Speech in Noise with Hearing Aids using Dual Microphones. Journal of the American Academy of Audiology 6 (6), 1995, 440-449.
31. Varibel Hearing Glasses. www.varibel.nl
32. Velichkovsky, B.M. and Hansen, J.P. New Technological Windows into Mind: There is More in Eyes and Brain for Human-Computer Interaction. In Proc. of CHI'96, ACM Press, 1996, 496-503.
33. Vertegaal, R. and Ding, Y. Explaining Effects of Eye Gaze on Mediated Group Conversations: Amount or synchronization? In Proceedings of CSCW 2002. New Orleans: ACM Press, 2002. 41-48.
34. Vertegaal, R., Slagter, R., van der Veer, G., and Nijholt, A. Eye Gaze Patterns in Conversations: There is More to Conversational Agents Than Meets the Eyes. In Proc. of CHI’01, ACM, 2001, 301-308.
35. Ware, C. and Mikaelian, H.H. An Evaluation of an Eye Tracker as a Device for Computer Input. In Proc. of CHI+GI '87, ACM Press, 1987, 183-188.
36. Wang, J., Zhai, S. and Su, H. Chinese Input with Keyboard and Eye-Tracking - An Anatomical Study. In Proceedings of ACM CHI’01 Conference on Human Factors in Computing Systems, 2001, 349-356.
37. Watanabe, J., Nii, H., Hashimoto, Y., and Inami, M. Visual Resonator: Interface for Interactive Cocktail Party Phenomenon. In Ext. Abstracts of CHI'06, ACM Press, 2006, 1505-1510.
38. Zhai, S., Morimoto, C., and Ihde, S. Manual and gaze input cascaded (MAGIC) pointing. In Proceedings of the ACM CHI 1999, 246-253.