Distributed Acoustic Conversation Shielding: An Application of a Smart Transducer Network

Yasuhiro Ono*, Joshua Lifton, Mark Feldmeier, Joseph A. Paradiso
Responsive Environments Group - MIT Media Laboratory
20 Ames Street, Cambridge, MA 02142 USA
(*Visiting Affiliate from Ricoh Company, Ltd, 16-1Shineicho, Tsuzukiku, Yokohama, 2240035, Japan)
+1-617-253-8988
[ono, lifton, geppetto, joep]@media.mit.edu

ABSTRACT
In this paper, we introduce distributed acoustic conversation shielding, a novel application of a transducer (sensors and speakers) network. This application protects the privacy of spontaneous conversations in a workplace by masking the participants' voices with sound from distributed loudspeakers that adapt to the dynamic location of the conversation vs. that of potential eavesdroppers. We demonstrate how the speakers collaborate with various sensors to produce masking sounds that satisfy the requirements of this application. An index of intelligibility, SNR (Signal-to-Noise Ratio) was used to evaluate the performance of our system. We suggest how the measured SNR can be used to adaptively servo the volume of the masking sounds.

Categories and Subject Descriptors
H5.5 [Sound and Music Computing]: User Interfaces. - Systems.

General Terms
Measurement, Design, Experimentation.

Keywords
Sensor Network, Conversation Shielding, Location-Awareness, Distributed Control, Sound Masking.

1. INTRODUCTION
Actuators such as speakers and lighting are commonly scattered throughout our living environments. As communication and sensing technologies have advanced towards the vision of ubiquitous computing [27], we have increasing opportunities to take advantage of such distributed actuators by using sensors that make them respond to the environment, thus increasing their utility and/or efficiency.

In this paper, we explore this idea by demonstrating an application of a distributed sensor and speaker network that provides a way for users to protect the privacy of conversations in the workplace, where many people meet and talk spontaneously. In offices, especially in increasingly-common open-space offices, violation of employees' privacy can often become an issue, as tertiary parties might overhear their conversations either intentionally or unintentionally. As face-to-face, spontaneous conversations among workers can result in a more productive and creative workplace, relieving the concern of being overheard is important. Existing solutions exploit products that mask conversations with background noise or other audio, which we term "acoustic conversation shielding". These, however, are products that output their audio from a single speaker with manual volume control, not adapting to the distribution of people and intrinsic background sound in the environment. This paper describes a distributed acoustic conversation shield encompassing a sound-masking system consisting of distributed speakers and sensors that automatically adjusts to its environment.

The structure of this paper is the following. Related works are shown in the remainder of this section. The detail of our approach, including a usage scenario, performance measurements, and an experimental deployment, are shown in section 2, then discussion and conclusions are presented in sections 3 and 4, respectively.

SANET'07, September 10, 2007, Montréal, Québec, Canada.

1.1 Related Works

1.1.1 Sensor Networks.
Technologies exploiting networked clusters of sensors have been developed to realize a broad range of applications. In particular, wireless sensor networks are expected to be deployed essentially everywhere (e.g., embedded in everyday objects to realize the dream of ubiquitous computing or unobtrusively collecting data on the environment), as the cost of the deployment will drop due to their denser integration and increasing energy efficiency [10, 29]. Today's prototypes of such wireless sensors are tools for building applications that explore the vision of ubiquitous sensor infrastructures [7].

Thus far, many sensor network applications have been proposed in wildlife and outdoor monitoring, showing scalability and low-power operation [15]. Other researchers have demonstrated workspace applications of sensor networks, such as conference room occupancy with motion sensors [5], while others have demonstrated home monitoring systems using wireless sensors [8]. These applications are basically aimed at monitoring what is happening or has happened in locations where it is costly or impractical for people to observe and collect data in person. On the other hand, our application exploits sensor devices around users, interpreting sensor measurements and automatically performing appropriate real-time actuation.

1.1.2 Location Awareness.
Indoor Location technology is one of the major needs of ubiquitous computing. A good overview of location technologies is found in [9]. Location accuracy has been improved with new technologies such as UWB (ultra wide band); for example, the Ubisense commercial system claims up to 15cm accuracy with

their active location tag and receivers set at the corners of a room [26], and UWB systems appropriate for integration into light-weight sensor networks are appearing [35]. Other recent approaches include adapting GSM [19] and power-line communication [20], which both exploit existing infrastructure.

Nonetheless, applications of localization technologies tend to lag and are still rather limited to established ideas such as location-aware guidance [1]. At a recent mobile computing conference, several location technology experts agreed that researchers in the field should focus more on applications, especially those that combine activity inference with location, instead of inventing a novel location technology [6]. Reflecting this opinion, we focus more on demonstrating what can be done through location-awareness.

1.1.3 Acoustic Conversation Shielding.

Sound-masking technologies are routinely used to reduce audio distraction and protect speech privacy in the workspace, such as in an open-plan office, reception area, or a meeting room. For example, conversations in meeting rooms can be protected partly by ceiling-mounted speakers that emit masking sounds. A recent commercial product [24] uses a set of speakers to emit recorded speech to mask a user's phone conversations.

However, the targets of known methods are limited to specific situations, such as telephone calls in a cubicle or discussions in a meeting room. Our target is spontaneous conversation that could happen at various places in a company, such as a corridor or a casual meeting space. Additionally, existing systems are self-contained boxes that are manually adjusted via a volume knob. Exploiting sensor networks and location awareness, we have set our goal to provide distributed, location-free acoustic conversation shielding in an automated and non-intrusive manner.

2. DISTRIBUTED ACOUSTIC SHIELDING

2.1 Usage Scenario

The usage scenario of our application is as follows. An office worker happened to meet one of his colleagues, a team member of a project, in the open space of their office area, and they started to chat about their project. He noticed that the content of the conversation was getting rather confidential to people outside their team. He pushes a button on his mobile device to trigger the acoustic conversation shielding application, at which point, various speakers surrounding them start to emit a masking sound to prevent others from overhearing the conversation. When the conversation is over, he pushes the button again to stop the masking. Even if he forgot to stop the speakers explicitly, the mobile device, embodied as a badge, detects the end of conversation with its microphone (and/or the dispersal of its participants via an IR proximity detector [33]) to turn the speakers off. Similarly, conversations could be autodetected by noting face-face proximity via IR along with alternating vocal cadence in the wearable mics.

To realize this scenario, the distributed audio system needs to thwart nearby listeners without disturbing people in the conversation or excessively irritating others in the area. Therefore, the volume of each speaker should be turned up only when it is located between the conversation source and someone nearby who is a potential listener. The masking sounds form a virtual barrier to acoustically isolate people involved in the conversation. Ideally, the volume of each speaker is automatically adjusted to the minimum needed to blind potential eavesdroppers. Although all speakers near potential listeners could be driven with masking sounds, the "cocktail party effect" [34] suggests that masking sources along the direction between eavesdroppers and the conversation are most effective.

2.2 Hardware

We used our Plug sensor network platform [13] as a networked speaker to emit the masking sounds. The Plug is designed as a ubiquitous sensing and actuation device for homes and offices (Figure 1 top and middle). It is modeled on a common electrical power strip, and it has various sensors, a wireless transceiver, and a speaker. In their role as power strips, Plugs have access to ample energy without batteries and already reside everywhere in homes and offices. A network of Plugs is an ideal candidate to bootstrap a backbone for ubiquitous computing, where Plugs communicate with wireless devices in the vicinity such as active badges and tags. Although not currently implemented in our prototype, power line communication (PLC) is an ideal network interface for the Plug.

Here we briefly introduce some of the Plug's functionalities especially related to this paper – more detail can be found in [13]. The Plug has a 32bit ARM7 microcontroller (Atmel AT91SAM764S) running at 48MHz, two programmable LEDs, a pushbutton, a speaker, a piezoelectric cantilever vibration sensor, a microphone, a phototransistor, a 2.4 GHz wireless transceiver (Chipcon CC2500), a current and voltage monitor on each outlet (the ARM can also turn each outlet on and off), and a USB 2.0 port. Here we also use an expansion board that contains a passive infrared (PIR) motion sensor and an SD memory card reader.

In our application, the cantilever vibration sensor and the PIR motion sensor are used to detect a person nearby. SD-cards are used to store audio clips that the Plug's speaker plays as masking sounds. The USB port is used to connect a Plug to a PC so that we can monitor the status of the Plug network on the PC's screen. Although the rich resources provided by the Plugs ease development, production platforms for such an audio masking application could be considerably streamlined and integrated into standard ceiling public address (PA) speaker deployment. For convenience, we call all nodes used by this application "Plugs."

To control the system remotely, we prepared a battery-operated mobile device (Figure 1 bottom), which bears the same functionality as the Plug (microcontroller, wireless transceiver, peripherals) without power functions and full sensing. This device provides a simple one-button user interface to control the masking audio. Its microphone can be used to detect conversation, provided users wear it over their shirt or jacket like a badge. We term this mobile device a "wearable controller" in this paper, and assume that users wear it when this application is running.

The Plug reads 8bit/8Hz PCM (Pulse-Code Modulation) audio data from the SD-card, and drives its speaker with PWM (Pulse-Width Modulation), as the ARM has no onboard DAC. We prepared three types of masking sounds that a user can launch with the wearable controller. One is a pre-recorded conversation, where the Plugs repeatedly play audio samples from the SD card. Another sound is a shuffled conversation, i.e., a continuous play of randomly-selected 640 millisecond slices of a pre-recorded conversation. The other sound is white noise, which is synthesized by the micro-controller. The volume of the speaker has three levels, which are named LOW, MEDIUM, and HIGH. In the Plug's firmware, the amplitude of the PWM modulation is set differently by each level; comparing amplitudes, volume HIGH is twice as large as volume MEDIUM and volume LOW is half as large as volume MEDIUM.

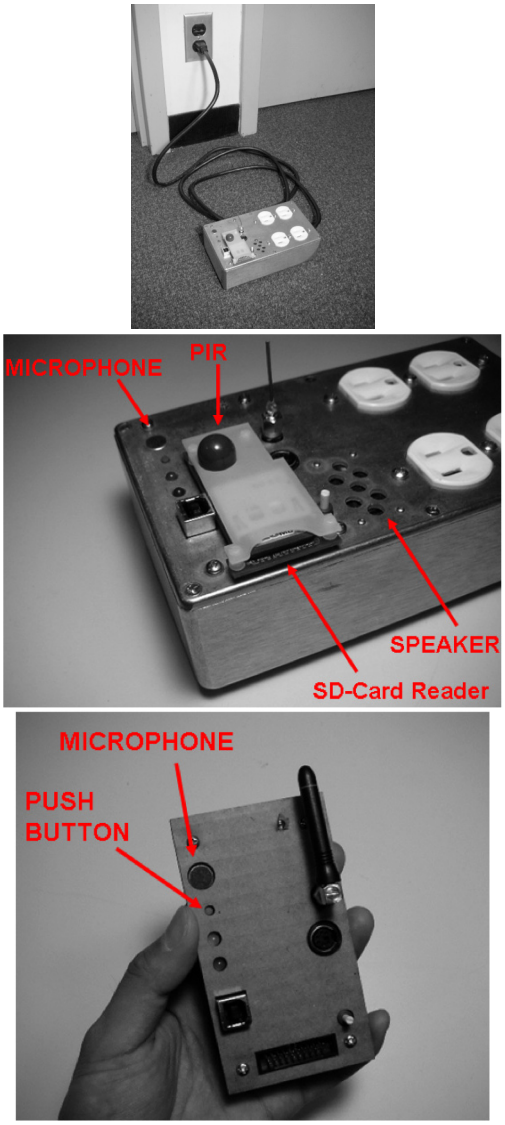

Figure 1. Plug sensor network platform prototyped as an electrical power-strip (top) and close-up of the sensor array (middle). The prototype wearable controller features a button to launch the masking operation. It also has a microphone to detect the user's conversation (bottom).

2.3 Masking Conversation

We assume that the Plugs make masking sounds when another Plug detects a person nearby, as depicted in Figure 2. This potential eavesdropper, whom we sometimes call a "listener" in this paper, hears both the conversation and the masking sound. In Figure 2, two people are talking while another person approaches. A person in the conversation wears a mobile controller that invokes the masking sound. The wearable controller is depicted as a triangle while two Plugs are depicted as rectangular. A Plug (left) emits a masking sound, when another Plug (right) detects the listener. The dark color of the Plug (right) means that the Plug has detected a person nearby. The intelligibility of the conversation to the listener is expected to be decreased by the masking sound.

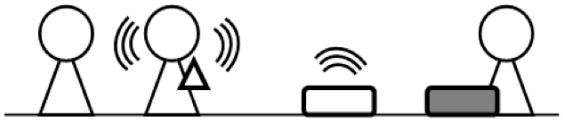

Figure 2. Diagram of conversation masking system. Two people (left) converse while a listener (right) approaches.

To evaluate the effectiveness and performance of masking, the intelligibility of the conversation to the listener should be considered. For that purpose, we introduce the notion of SNR (Signal-to-Noise Ratio), which is often used as an index of intelligibility in the research of auditory perception [16, 2, 4, 30, 31]. SNR is the ratio of the sound power of the target speech to that of noise. The masking sounds increase noise energy. A pioneering study of masking sound showed that intelligibility decreases monotonically as SNR decreases [16].

It has been suggested that masking sounds having characteristics similar to the target speech decrease intelligibility effectively; speech by the target person masks better than the speech by a different person, a different sex, or noise [2, 4, 30]. For example, the score of an intelligibility test in Brungart's study [4, 30] is decreased from 80% to 40% as SNR was decreased from 6bB to 0dB when they used speech by the same person as a masking sound. In our application, we might assume that users record their speech in advance so that Plugs can use snippets of their speech or exploit a model of their vocal characteristics as a masking sound when the application is invoked.

As detailed below, we conducted an experiment to estimate SNR at the position of the listener in an experimental setting where the volume of the masking sound was set at various levels and the distances between the listener and the conversation was changed. We used two streams of audio measurements; one from the wearable controller's microphone and the other from the microphone of a Plug that was put at the listener's position. We used a high-quality speaker driven by a PC to mimic the conversation, and the wearable controller was put close to this speaker. We placed a Plug at 2, 3, 4, and 5 meters away from the speaker (assuming that the listener is in this position), and put another Plug in the middle between the conversation and the listener to provide a masking sound. We used a speech corpus consisting of short sentences recorded by three males and three females [17] for both the target conversation played by the PC speaker and the masking sound. The sentences are called "Harvard psychoacoustic sentences," which were developed for subjective measurements of speech [11]. The PC speaker and the Plug repeatedly played excerpts from the speech corpus. The PC put a short pause of around 5 seconds between each "conversation" sentence. With a commercial sound level meter located 30 cm from the acoustic source, the peak loudness of the speaker was around 75-85 and 70-80 dB SPL (A) for the PC speaker and the Plug's speaker with "MEDIUM" volume, respectively. The loudness of the Plug's speaker was decreased by around 3 dB SPL and increased by around 3 dB SPL at "LOW" volume and at "HIGH" volume, respectively. SNR was calculated with 90-second recordings of microphone measurements when both speakers were turned on.

Figure 3 shows the sound power of the wearable controller (top) and the Plug representing the listener (bottom) when the distance was 5m and the volume of the Plug was LOW. The sound power was calculated every 192 milliseconds in each microcontroller, at which 8bit/8Hz microphone measurements were used. As shown in the figure, the wearable controller's micro-

phone is saturated in the presence of speech. The presence and absence of conversational speech is thus easily detected by setting a threshold on sound power at the wearable controller.

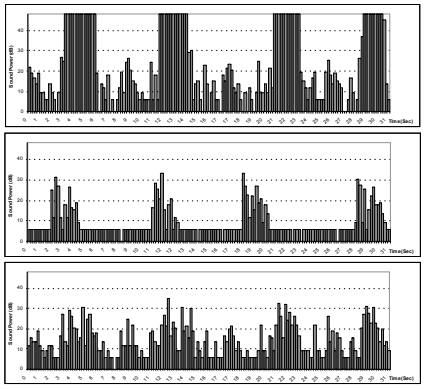

Figure 3. Sound power measured by a wearable controller at the position of a conversation (top). A Plug was placed in the middle between the conversation and the listener to provide a masking sound. The middle and bottom plots show the sound power measured by a Plug at the position of a listener without (middle) and with masking (bottom). The distance between the conversation and the listener was 5m and the volume of the masking sound in the bottom plot was LOW.

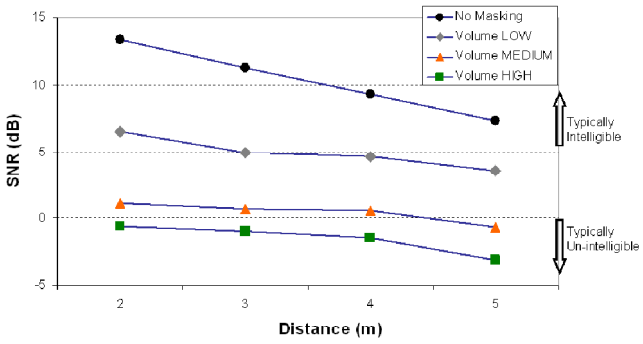

Figure 4. Calculated Signal-Nose Radio (SNR) when the distance between the conversation and the listener is 2, 3, 4 and 5m and the volume of the masking sound is LOW, MEDIUM, and HIGH. SNR without the masking sound is also presented. 90-second sound power measurements at the wearable controller and a Plug are used to calculate SNR. SNR is seen to decrease with distance and the volume of the masking sound.

To calculate SNR, we need two measurements of sound power - the target speech and induced noise - at the listener's position. Noise power was calculated as a time average of the sound power during the absence of the conversational speech. The power of the target speech was calculated by subtracting the noise power from a time average of the sound power during the presence of the speech in the conversation. As seen in Figure 3, the masking sound dominated the room's ambient noise sources (middle plot).

Figure 4 shows the calculated SNR when the distance is 2, 3, 4 and 5m and the volume of the masking sound is LOW, MEDIUM, and HIGH. SNR with no masking sound is also shown. SNR decreases as the distance or the volume of the masking sound increases. If we apply the psychoacoustic study that claims intelligibility drops when SNR decreases from 6dB to 0dB [4, 30], this result might be interpreted as follows: The masking sound decreased the intelligibility especially when the volume was MEDIUM or HIGH, while the listener could rather understand the conversation when no masking sounds were presented.

It is worth mentioning that the psychoacoustic study also suggests that the decrease of the intelligibility was not observed when SNR decreased beyond 0dB when speech was used as a masking sound. If we apply this principle, masking sounds with HIGH volume are more than needed for masking purposes across these distances, and MEDIUM volume was sufficient out to 5m, for example. As emitting redundant sound power into the environment is undesirable, it behooves us to keep the volume of the masking sound limited.

Notice that these interpretations represent our initial speculation of the intelligibility. Rigorous subjective evaluations are needed to precisely estimate intelligibility from SNR.

2.4 Location Awareness

We assume that Plugs and wearable controllers know the two dimensional (x,y) coordinate of their location in the environment. To test our application in a location-aware setting, the Plugs read their assigned locations from the SD-card at the time of booting. We used these pre-fixed coordinates in the deployment experiment presented later in this paper. RSSI (Radio Signal Strength Indicator) based location estimation [3] could be implemented into Plugs and wearable controllers, assuming that a set of anchor plugs have pre-fixed coordinates. An RSSI-based location algorithm is implemented on the family of wireless transceivers that Plug uses [25].

As eavesdroppers may not be wearing badge transmitters, listener locations are roughly estimated in our system by the Plugs' vibration and PIR motion sensors – such sensors, in a sufficiently dense deployment, have been shown to be able to track occupants through a building provided enough state is retained [28].

2.5 Control Scheme

Decentralized control that exploits local computation is a natural choice for a distributed system, since it does not depend on either a central controller or a central storage, which could become a bottleneck to the system's scalability and response. Therefore, we control the speaker of the Plugs in a decentralized manner, letting each Plug manage its own speaker for faster response and easy expandability when users introduce additional Plugs into the system. To separate control code from lower-level routines (such as communicating with neighbor Plugs) in the firmware, we prepared "neighbor caches" [14], a table consisting of the latest sensor measurements of the neighbor Plugs. The response of a Plug's speaker is developed by consulting this neighbor cache to account for the state of other Plugs in the neighborhood. We will illustrate this later in an example of the control code.

Table 1 shows a Plug's neighbor cache as prepared for each nearby Plug. It includes sensor measurements (microphone, PIR, and vibration) and the status of the speaker (type and volume of generated sound). It also keeps the (x,y) coordinate, an address that is unique among neighboring Plugs, and the RSSI and time stamp taken when receiving the last radio packet. Plugs update their neighbor cache when they receive a packet from another Plug. Every 192 milliseconds, each Plug calculates the averaged sound power from 8bit/8Hz microphone measurements. After obtaining these values eight times, the Plug transmits a packet that contains the sequence of sound powers with PIR and vibration sensor measurements averaged over the last 1.5 seconds. The packet also contains the coordinate, the status of the speaker, and

the node address. When a Plug receives this packet, it updates the values of the neighbor cache corresponding to the transmitting node's address.

Table 1. Neighbor cache includes sensor measurements and status of the speaker of each Plug in the neighborhood.

Item	*Description*
Address	Unique ID among neighboring devices
Microphone	Averaged sound power
Passive IR (PIR)	Is PIR activated?
Vibration	Is vibration detected?
Speaker Volume	Volume of the Speaker
Speaker Sound	Type of Speaker Sound
Location	(x,y) coordinate
RSSI	Radio Signal Strength Indicator
Time Stamp	Time when the last packet was received

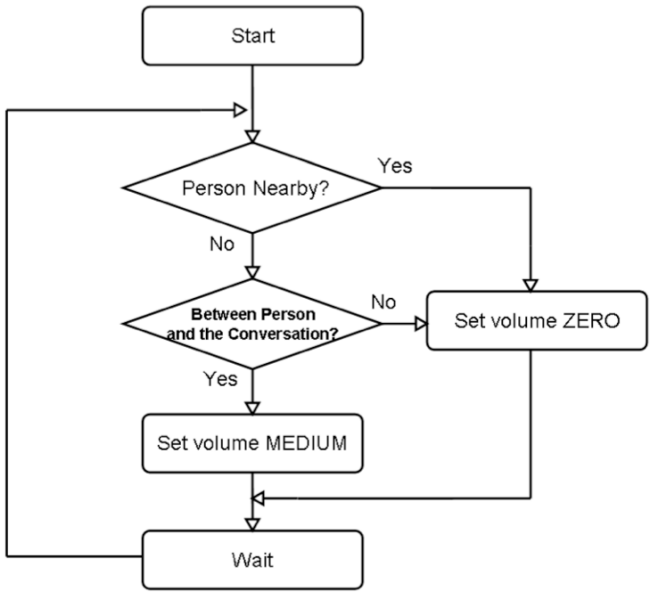

Figure 5. Flowchart of audio control code for a deployment experiment. The code was designed so that Plugs make a dynamic sound barrier between the conversation and listeners.

3. Deployment and Testing

Twelve Plugs were deployed in our lab space on the 3rd floor of MIT Media Lab to test this application. We connected another Plug to a PC to monitor the status of the 12 participating Plugs. Figure 6 is a snapshot from the monitoring software running on the PC with additional manual annotation for explanation, where each Plug is drawn as a rectangle and the wearable controller is drawn as a triangle, as in Figure 2. The additional symbols are explained later. Plugs are deployed at positions shown in the figure, where each Plug is about 2 meters away from its neighbors. These locations were cached as (x,y) coordinates in each Plug and in the mobile controller as we explained earlier. The filled rectangle means the Plug detected a person nearby with the PIR or vibration sensor, while the horizontal lines above a rectangle mean that the Plug is making a masking sound. The screenshot was captured after a user activated the masking system with the wearable controller and one of the Plugs detected a listener who was at the position labeled "C". At the time of the screenshot, 3 Plugs were making a masking sound to shield the users' conversation.

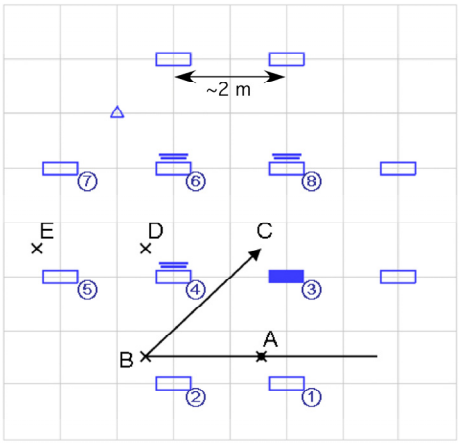

Figure 6. Screenshot from the monitoring software on a PC, which visualizes the status of 12 Plugs (rectangles) and a mobile controller (triangle). Each Plug is about 2 meters away from the nearest Plug. A filled rectangle means the Plug detected a person nearby with the active PIR or vibration sensor, while the horizontal lines above a rectangle mean that the Plug is making a masking sound. This screenshot was captured after a user turned on masking sounds with a wearable controller, when one of the Plugs detected a person who was at the position labeled "C". Three Plugs are making a masking sound and shielding the user's conversation.

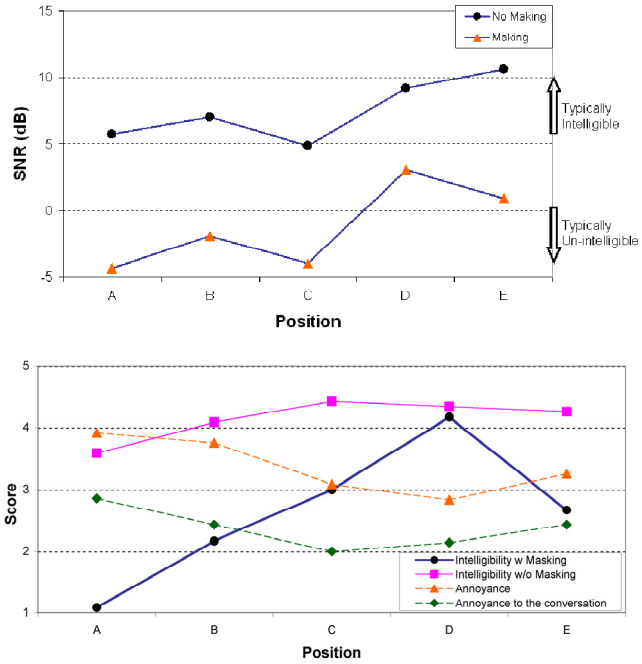

Figure 7. *Top:* Calculated Signal-Noise Radio (SNR) at the positions labeled A, B, C, D, and E. SNRs without masking sounds are also presented. 90 second recordings of the sound power measurements at the wearable controller and at a Plug are used to calculate SNR. *Bottom:* Results of human subject tests quantifying intelligibility with and without masking (solid curves), and annoyance to people at the eavesdropper's location and distraction to the conversing individuals (dotted curves) as a function of listener position.

The control code managing the speaker of each Plug was explained with the flowchart shown in Figure 5. This procedure was designed so that Plugs make a sound barrier between the conversation and listeners. The routine begins by checking whether the Plug has detected an active PIR or a vibration sensor, which is interpreted as someone nearby. If nobody is detected, it checks whether (i) there are any neighbor Plugs that have detected a person nearby and (ii) the Plug is located between the neighbor and the wearable controller. The code then estimates the relative positions of Plugs and wearable controller from the coordinate values in the neighbor cache. If both (i) and (ii) are true, the code turns up the volume to MEDIUM so that the Plug becomes a part of the sound barrier. Otherwise, the code keeps the volume zero. This process is repeatedly invoked to reflect any change in the environment. We confirmed that the sound barrier is adjusted as a person walked through the environment, along a path labeled A, B, C, D, and E in Figure 6 for example; as seen in the figure, Plugs in the appropriate positions emitted a masking sound.

To evaluate whether the sound barrier successfully masked the conversation, we calculated SNR at the positions labeled A, B, C, D, and E in Figure 6. The experimental setting is the same as described earlier. We used a high-quality PC speaker to mimic the conversation and put another Plug at each listener's position for the SNR measurement. The same speech corpus was used for the content of the conversation and the masking sound. A wearable controller was placed beside the PC speaker to detect the presence of speech in the conversation. As before, we estimated SNR from two streams of microphone measurements for 90 seconds; one from the wearable controller's microphone and the other from the microphone of the Plug at the listener's position.

The results are shown at the top of Figure 7. The masking sounds decreased SNR by 5-10 dB at each location. If we apply the psychoacoustic study that we have adopted [4, 30] saying that intelligibility drops when SNR is decreased from 6dB to 0dB and speech is unintelligible below 0dB, the result could be interpreted as follows. At the positions D and E, which are closer to the conversation, SNR was between 0 dB and 6 dB, meaning the intelligibility was decreased but it could be decreased more if the volume of the masking were increased. At the positions A, B, and C, which are more than 5m distant from the conversation, SNR dropped below 0 dB, meaning the masking sound decreased intelligibility sufficiently and may be louder than needed.

In order to test this indication in more detail, we recorded audio of a conversation (again extracted from the Coordinated Response Measure (CRM) speech corpus of [4,30]) with and without masking, as heard at each location of Figure 6. We then played this back through earbuds for seven subjects, who rated the conversation's intelligibility on a scale of 1-5 at each position. They also rated their annoyance at the masking sound for audio recorded at the positions of the eavesdropper and the conversation. Evaluating this system with sound recorded at each position isn't completely faithful, as, for example, it eliminates any directional cues that the listener picks up by moving his or her head, etc. On the other hand, this technique guarantees a stable acoustic environment, as we were unable to secure a test space with consistent background noise.

Results are shown at the bottom of Figure 7 (averaged across all users), where one can see the intelligibility of the masked conversation steadily increase as the user approaches the conversation. The speech was deemed as understandable as the unmasked audio when one approaches to within circa 3 meters, due to the fewer number of masking speakers activated and louder primary sound level (note that, although points D and E were roughly equidistant from the conversation, users seemed to rank position E less understandable, probably because of adaptation effects – as all they experienced the audio stream progressing from points E to A, their ears became better accustomed to the quality of speech after point E). Positions further from the conversation than point C (roughly 5 meters away) were rated less than half intelligible, which is somewhat in accordance with the SNR predictions at the top of Figure 7.

The users also related a subjective "level of annoyance or distraction" at each potential eavesdropper position (here assuming that the "eavesdroppers" are actually other employees hard at work), as well as the amount of distraction from the masking sounds present at the conversation. Figure 6 (bottom) indicates that annoyance drops a bit as the eavesdroppers approach the conversation, again because there are fewer speakers making masking sound. The amount of distraction to the conversing partners is consistently rated below midpoint and is always well below the annoyance to the eavesdropper.

Figure 8 illustrates the system in operation, where the plug speakers are seen to automatically switch as the user walks the course of Figure 6. This system is essentially open loop – the masking audio played at each node is determined solely by activity detected by the motion sensors in the network and the relative position of eavesdropper vs. conversation. Note that, as illustrated in Figure 5, only one level of masking audio can currently be selected. We feel that the performance of the system could improve significantly if the masking audio at each speaker could be continuously varied under distributed audio feedback control.

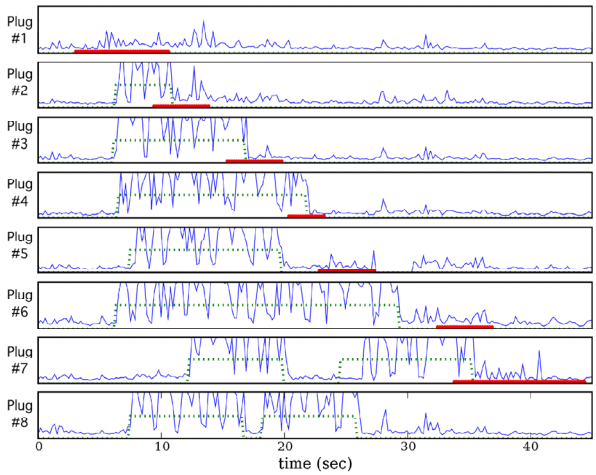

Figure 8. Microphone amplitude for Plugs 1-8 as the listener walks from A-B-C-D-E (see Fig. 6), showing dynamic response of masking audio at each plug to changes in the listener's location. The microphone signal (solid curve) tends to saturate when the plug's speaker is activated (indicated by the dotted line). Motion sensor detection of the listener is indicated by bold red bars on the horizontal axes.

4. DISCUSSION

4.1 Feedback Speaker Control

While we observed that the masking sounds decreased SNR, an index of the intelligibility to a listener in our experimental settings, regulating the continuous volume of the masking sounds dynamically to an appropriate level could improve performance.

Here the system would adjust the volume of the speakers under feedback control while targeting a quantity measured by distributed microphones. Assuming that SNR to the listener could be approximated by SNR calculated at a quiet Plug near the listener together with sound levels at the conversers' wearable controllers, the control code for the masking sound can use this estimated SNR for adjusting the volume. Designing and implementing this feedback control loop is a challenge for future work.

We could also measure the induced masking sound level at position of the conversation with the wearable controller. Since background noise heard at the conversation can quantify how much the conversers are perturbed by the masking sounds, it could be used as another quantity for adjusting the volume. Furthermore, a multivariable problem is posed when considering several simultaneous conversations with multiple eavesdroppers. This problem can be formulated as adjusting the masking amplitude at each plug to optimally shield the conversationalists from the eavesdroppers while minimizing or bounding the masking-induced distraction to the conversing people and any noise-related disturbance to others. This is also a topic for future work.

To calculate SNR, we used only the received sound power in our experiments, setting a threshold to separate speech and silence at the wearable controller. The process of segmenting speech and silence is often called voice activity detection. A recently developed wearable badge from our team [18] exploits an analog filter with several frequency bands selected for detecting speech. Such audio processing could be used to better estimate SNR.

We could also employ various voice recognition technologies to quantify intelligibility to the eavesdropper. Similarly, the wearable controller could transmit very short grains of compressed conversational audio that could be correlated with audio received near potential listeners to more precisely quantify signal-to-noise. Such approaches may pose a privacy concern, because our Plug can be thought to be similar to a "bug," an eavesdropping device that intentionally invades privacy. Thus, we should be careful neither to store nor transmit signals of sufficient duration or quality to discern the content of the conversation. Fortunately, this application only requires whether, or how much, the conversation is being leaked, not the content of the conversation.

This system must also be socially acceptable – when activated, the system must fade on gradually without a shocking, abrupt transient. Similarly, other sources of masking noise, such as background music, could be adopted. The small speakers on our present plug hardware exhibit limited quality – superior performance will be attained with speakers of higher fidelity.

4.2 Location Technology

The relative locations of the Plugs, conversers (wearable controller), and the eavesdroppers need to be roughly estimated for this system. Although we did not evaluate our system with location-aware operation beyond motion sensors for eavesdropper detection, we mentioned an established RSSI approach [3], which uses several fixed beacons as location references. A deficiency of this method, in addition to its inaccuracy, is that the reference nodes need to know their location. Considering that our application requires only coarse, relative neighborhood positions, especially within the range of audible acoustic signals, a good alternative approach is acoustic-based localization, such as ToA (Time-of-Arrival) and AoA (Angle-of-Arrival). ToA localization with ultrasound has been investigated for many years [21]. Scott el. al. showed a ToA localization approach to detect human sounds such as finger clicks for 3D user interfaces [23]. Calibration of a microphone and speaker network with the ToA of audible sound was presented in [22], although the range is limited to 2-3 meters due to the attenuation of the audible sound. Girod et. al. showed an acoustic AoA estimation with a 4-channel microphone array, and their prototype obtained 1.5 degree average orientation error in their outdoor experiment with using a chirp sound [32]. Another alternative is to exploit environmental signals to estimate node location. Wren et. al. [28] showed that data from simple motion detectors can statistically derive the spatial arrangement of the sensors. In our application, natural sonic transients, such as door slams or footsteps, could also be exploited to determine relative node positions [12].

4.3 Network Architecture

In our current prototype, the only network interface on the Plug is a generic radio, as often employed in wireless sensor networks [10, 29]. This is suitable for communication with low-power wireless devices, such as the wearable controller in our application.

In addition to the radio, employing higher-bandwidth power-line communication would be beneficial for transferring large quantities of data, such as digital content. For example, in our application, audio data does not need to be stored in an SD-card in advance if we can transfer audio samples from a central server on demand. Both communication channels could be useful in the network architecture of the speaker and sensor network that we propose.

5. CONCLUSION

In this paper, we introduced a novel application of a transducer network: distributed acoustic conversation shielding. This application protects the privacy of conversations that happen spontaneously in a workplace by masking the voices with sound from distributed loudspeakers. We demonstrated how distributed speakers with various sensors collaboratively generate masking sounds to satisfy the requirements of this application. We argue that the masking performance can be measured as SNR (Signal-to-Noise Ratio), which is calculated by the network of microphones. The results of our experiments suggest that it is beneficial to introduce feedback control into the application, where the volumes of the masking sounds are continuously controlled by using distributed microphone measurements. Our study suggests that there are ample opportunities to advance the proposed application by integrating various fields of research, including psychoacoustics, sensor networks, control theory, and location awareness.

6. ACKNOWLEDGMENTS

We would like to thank the Things That Think consortium and all the sponsors of the MIT Media Lab for their support. We also thank Atsuo Shimada and Soichiro Iga at Ricoh for their support of this work.

7. REFERENCES

[1] Abowd, G.D., Atkeson, C.G., Hong, J., Long, S., Kooper, R., and Pinkerton, M. "Cyberguide: a mobile context-aware tour guide." *Wireless Networks*, 3(5), (Oct. 1997), 421-433.

[2] Assmann P.F., Summerfield A.Q., "The perception of speech under adverse conditions." In S. Greenberg, W.A. Ainsworth, A.N. Popper and R. Fay (Eds.) *Speech Processing in the Auditory System*. Springer-Verlag, New York. 2004, pp. 231-308.

[3] Bahl, P., and Padmanabhan, V., "RADAR: An In-Building RF-based User Location and Tracking System." In *Proc. IEEE INFOCOM* (Tel-Aviv, Israel, Mar. 2000), Vol. 2, pp. 775-784.

[4] Brungart D. S., "Informational and Energetic Masking Effects in Multitalker Speech Perception," In Divenyi, P. (Eds.), *Speech Separation by Humans and Machines*, Kluwer Academic Publishers, 2005, pp. 261-267.

[5] Conner, W.S., Chhabra, J., Yarvis, M., and Krishnamurthy, L. "Experimental evaluation of topology control and synchronization for in-building sensor network applications." *Mobile Networks and Applications*. Vol. 10, Issue 4, 2005, pp. 545-562.

[6] Ebling, M.R., "HotMobile 2006: Mobile Computing Practitioners Interact," *IEEE Pervasive Computing*, Volume 5, Issue 4, Oct.-Dec., pp. 102-105, 2006.

[7] Estrin, D., Govindan, G., Heidemann, J., and Kumar, S., "Next century challenges: Scalable coordination in sensor networks." In *Mobile Computing and Networking*, pages 263–270, 1999.

[8] Fogarty, J., Au, C., and Hudson, S.E. "Sensing from the basement: a feasibility study of unobtrusive and low-cost home activity recognition." In *Proc. of the 19th Annual ACM Symposium on User interface Software and Technology UIST 2006* (Montreux, Switzerland, October 15-18, 2006), pages 91-100.

[9] Hightower, J., and Borriello, G., "Location Systems for Ubiquitous Computing," *Computer*, 34(8), Aug. 2001, pp. 57–66.

[10] Hill. J., Szewczyk. R., Woo. A., Hollar, S., Culler, D., and Pister., K., "System architecture directions for network sensors." In *Architectural Support for Programming Languages and Operating Systems*, 2000, pp. 93-104.

[11] IEEE Recommended Practice for Speech Quality Measurements, *IEEE Transactions on Audio and Electroacoustics*, Vol. 17, pp. 227-46, 1969. Also found on: http://www.cs.columbia.edu/~hgs/audio/harvard.html

[12] Kim, D.S., *Sensor Network Localization Based on Natural Phenomena*, M.Eng. Thesis, MIT EECS & Media Lab, 2006.

[13] Lifton, J., Feldmeier, M., Ono, Y., Lewis, C., and Paradiso, J.A., "A Platform for Ubiquitous Sensor Deployment in Occupational and Domestic Environments," *International Conference on Information Processing in Sensor Networks (IPSN 07)*, Cambridge, MA, 25-27 April 2007, pp. 119-127.

[14] Lifton, J., Seetharam, D., Broxton, M., and Paradiso, J., "Pushpin Computing System Overview: a Platform for Distributed, Embedded, Ubiquitous Sensor Networks," *Proceedings of the First International Conference on Pervasive Computing*, pp.139-151, August 26-28, 2002.

[15] Mainwaring, A., Polastre, J., Szewczyk, R., Culler, D., Anderson, J., "Wireless Sensor Networks for Habitat Monitoring," *WSNA '02*, Sept. 2002, Atlanta, GA, USA, pp. 88-97.

[16] Miller G. A., "The masking of speech", *Psychological Bulletin* 44(2), pp. 105-129, 1947.

[17] *A noisy speech corpus (NOISEUS)*, http://www.utdallas.edu/~loizou/speech/noizeus/

[18] Olguín, D.O., Paradiso, J., and Pentland, A.S., "Wearable Communicator Badge: Designing a New Platform for Revealing Organizational Dynamics." *IEEE 10th Intl. Symposium on Wearable Computing (Student Colloquium Proceedings)*. Montreaux, Switzerland. October 11-14, 2006, pp. 4-6.

[19] Otsason. V., Varshavsky. A., LaMarca, A., Eyal de Lara, "Accurate GSM Indoor Localization," *Proceedings of the Seventh International Conference on Ubiquitous Computing (UbiComp2005)*, pp. 141-158, Tokyo, Japan, 2005.

[20] Patel, S.N., Troug, K.N., and Abowd, G.D., "PLP: A Practical Sub-Room-Level Indoor Location System for Domestic Use," *Proceedings of the 8th International Conference on Ubiquitous Computing (UbiComp2006)*, pp. 441-458, Orange County, USA.

[21] Priyantha N.B., Chakraborty, A., Balakrishnan, H., "The Cricket location-support system," *Proceedings of the 6th annual international conference on Mobile computing and networking*, pp. 32-43, August 6-11, 2000, Boston, Massachusetts, United States.

[22] Raykar, V.C., Kozintsev, I., and Lienhart, R., "Position calibration of audio sensors and actuators in a distributed computing platform." In *Proceedings of the Eleventh ACM international Conference on Multimedia - MULTIMEDIA '03* (Berkeley, CA, USA, November 02 - 08, 2003). ACM Press, New York, NY, 2003, pp. 572-581.

[23] Scott, J., and Dragovic, B., "Audio Location: Accurate Low-Cost Location Sensing." *Proc. of The Third International Conference on Pervasive Computing*, pp. 1-18, 2005.

[24] *Babble®*, Sonare Technologies, http://www.sonaretechnologies.com

[25] Taubenheim, D., Kyperountas, S., and Correal, N., *Distributed Radiolocation Hardware Core for IEEE 802.15.4*, http://www.chipcon.com/files/Radiolocation_Engine_WP.pdf

[26] *Ubisense*, http://www.ubisense.net (2007).

[27] Weiser, M., "The computer for the 21st century." *Scientific American*, 265(3), pp. 94–104, 1991.

[28] Wren C.R., and Rao, S.G., "Self-configuring, lightweight sensor networks for ubiquitous computing." In *The Fifth International Conference on Ubiquitous Computing: Adjunct Proceedings*, 2003, pp. 205–6, also MERL Technical Report TR2003-24.

[29] Crossbow Technology. http://www.xbow.com

[30] Brungart, D. S., Simpson, D. B., Ericson, M. A., and Scott, K. R., "Informational and energetic masking effects in the perception of multiple simultaneous talkers," *J. Acoust. Soc. Am.* 110(5), pp. 2527-2538, 2001.

[31] Brungart, D. S., "Informational and energetic masking effects in the perception of two simultaneous talkers," *J. Acoust. Soc. Am.* 109(3), pp. 1101-1109, 2001.

[32] Girod, L., Lukac, M., Trifa, V., Estrin, D., "The design and implementation of a self-calibrating distributed acoustic sensing platform," *Sensys 06*, ACM, 2006, pp. 71-84.

[33] Laibowitz, M., et al., "A Sensor Network for Social Dynamics," in the *Proc. of the Fifth Int. Conf. on Information Processing in Sensor Networks (IPSN 06),* Nashville, TN, April 19-21, 2006, pp. 483-491.

[34] Arons, B., "A Review of The Cocktail Party Effect," *Journal of the American Voice I/O Society* 12, July 1992, pp 35-50.

[35] K. Mizugaki, et al, "Accurate Wireless Location/Communication System With 22-cm Error Using UWB-IR," in *Proc. of the 2007 IEEE Radio and Wireless Symposium*, pp. 455-458.