Compromising Speech Privacy under Continuous Masking in Personal Spaces

S Abhishek Anand
University of Alabama at Birmingham
Email: anandab@uab.edu

Payton Walker
University of Alabama at Birmingham
Email: prw0007@uab.edu

Nitesh Saxena
University of Alabama at Birmingham
Email: saxena@uab.edu

***Abstract*—This paper explores the effectiveness of common sound masking solutions deployed for preserving speech privacy in workplace environment such as hospitals, financial institutions, lawyers offices, nursing homes and government buildings. With the increased awareness about personal privacy among the general population, we set out to examine the effectiveness of current speech privacy preserving tools. We seek to determine if the general approach used by the current masking mechanisms is adequate to provide the level of privacy desired from these solutions. In addition, we also seek to investigate preservation of speech privacy in the face of ubiquitous and less conspicuous devices like smartphones that possess the capability of sound recording with inbuilt noise cancellation technology.**

Our approach in this paper is to expose the vulnerability in sound masking technology in scenarios that require preserving privacy in personal spaces. We use human listeners to attack speech privacy under sound masking where we aim to identify spoken words eavesdropped under different scenarios. We also test currently available speech recognition tools to assess their performance at decoding speech in noisy environment. Our results indicate that pink noise, the commonly used technology to provide speech privacy for use in personal space, is ineffective against a dedicated eavesdropping adversary that uses commonplace devices such as smartphones to record the speech and noise reduction tools to counteract sound masking.

I. Introduction

Speech privacy can be described as the inability of an unintentional listener to understand another person's conversation [2]. A study conducted by the Center for the Built Environment (CBE) at UC Berkeley in 2003 [22], showed that almost 72% of office workers do not consider their workplace as a safe environment for speech privacy. The current workplace designs include *open office space* that can be defined as workspace where the perimeter boundaries do not go to the ceiling and *private office space* where the location has four walls extending to the ceiling and a door [6].

In real-world scenarios, speech privacy can be compromised if an unintended listener can overhear a confidential conversation. As an example, a patient's private details like medical prescriptions could be discussed between a doctor and a nurse in a hospital ward. If the conversation is overheard by other patients, doctors, or medical staff present in the ward, then the speech privacy of that conversation would be compromised.

Open office spaces have been heavily promoted at workplaces by several companies who believe that such design facilitates a better flow of communication and interaction within the workforce; thereby improving overall performance and satisfaction of the team [4]. However there exist several studies that point out the detrimental effect of open office spaces on speech privacy due to lack of inhibitive measures against noise. A survey by Karleela-Tuomaala et al. [17] showed that there was a significant loss of speech privacy felt by people when they were moved from private office rooms to open office space. Loss of speech privacy may lead to less communication as people may refrain from carrying out work related sensitive communication that could potentially be eavesdropped in the surrounding area [5].

To combat the dangers facing speech privacy at workplace and offices, three principles commonly referred to as "The ABC's", are used to maintain the desired privacy levels [18], [25]. "A" stands for absorb, "B" stands for block and "C" stands for cover. Each principle is implemented individually, but their combined contributions aid in achieving the desired level of privacy. Sound masking is an implementation of the "Cover" principle wherein a masking sound in the desired spectrum is continuously generated in the background at the required location to hide speech. Such sound masking setups consist of loudspeakers called emitters, and a control panel that regulates the volume of the produced sound [18]. The emitters can be mounted on walls or ceiling and can be invisible to unsuspecting eavesdroppers. They can be used in open office spaces like call centers, research areas, hospitals and other medical facilities, as well as in closed office spaces like corporate boardrooms, lawyers' offices and financial institutions.

Many commercial solutions are present in the market and deployed in real-life settings that aim to provide speech privacy in an individual space through sound masking. For example, Sonet Qt® from Cambridge Sound Management is geared towards reducing distractions in an individual's personal space. AtlasIED offers the sound masking system UL2043 and has selectable white and pink analog sources for masking sounds. Speech Privacy Systems provides different masking solutions for cubicles, private offices, medical facilities and call centers.

Our Contributions: In this paper, we show the ineffectiveness of sound masking in maintaining speech privacy in different workplace settings. We use human eavesdropper to decode the speech samples which are recorded under multiple loudness levels of masking sound consisting of pink noise (preferred for speech masking [13]) and in different workplace settings. We also make use of common speech recognition tools such as

978-1-7281-3265-5/19/$31.00 ©2019 IEEE

Google Cloud Speech API that can be used to extract speech from the eavesdropped samples. We show that a malicious eavesdropper can compromise speech privacy by using low cost, off the shelf devices such as PC microphones and smartphones. To provide additional capability to the eavesdropper, we also use noise reduction using spectral subtraction to reduce the masking sound from the recorded samples making it more intelligible for the eavesdropper.

1) ***Compromising speech privacy through human eavesdropper:*** We designed a study on the Amazon Turk platform to decipher speech samples recorded in three workplace designs: *open office* (Figure 1a), *closed office* (Figure 1b) and *hybrid/semi-private office*. We show that in the *open office* design, the best attack accuracies at decoding speech are obtained (over 90%(60%) for male(female) speakers at 65dB; approximately 70%(40%) for male(female) speakers at 55dB) when speech is comparable or louder (+10dB) than the masking sound.
   The *closed office* design seemed to be effective at preserving speech privacy when the speech volume is comparable to the masking sound volume. The results showed that the speech was incomprehensible to listeners (17% for male speaker). However the masking proved ineffective when the speech was louder than the masking sound (90% and 100% for male and female speakers 10dB louder than masking sound at 55dB).
2) ***Compromising speech privacy through speech recognition tools:*** We test the eavesdropped speech samples against speech recognition tools to determine if it is possible to perform an automated attack where human eavesdropping is not required. We also test the performance of these tools against human listeners to see if they are better at speech recognition under noisy conditions. We show that automatic speech recognition (ASR) tools are not as competent in extracting speech from the noisy samples as human listeners, especially when SNR approaches 1. In the *open office* scenario, ASR achieves an accuracy of 92% and 62% with loud male and female voices (65dB) in the presence of noise at 55dB. These accuracies are however degraded to 85% and 31% respectively, after noise reduction. ASR performs worse in the *closed office* and *semi-closed office* scenarios where the decoding accuracies drop to 0% with an SNR of 1 or lower.
3) ***Effect of noise filtering:*** We show that noise filtering techniques that use spectral subtraction can be used to improve upon the decoding accuracies by making the recorded samples more intelligible for human listeners. For the *open office* scenario, we report an increase in maximum possible decoding accuracy for female speakers (65dB) at a 75dB noise level (23% $\rightarrow$ 38%) and for male speakers (55dB) at a 55dB noise level (69% $\rightarrow$ 77%). For the *closed office* design, the average accuracies increased for male speakers (55dB) in a noise level environment of 55dB from 32% to 58% while for male speakers at 65dB and at the same noise level, it increased from 30% to 43%. The maximum accuracies are also boosted for female speakers (65dB) with a noise level of 65 dB (14% $\rightarrow$ 28%), and with a noise level of 55dB (86% $\rightarrow$ 100%). A similar trend was observed in the *semi-closed office* scenario for the same speakers and noise levels.

Our work demonstrates that any deployed solution that uses psuedorandom noise for preserving speech privacy may not be effective. Since a masking sound level is expected to be around 45-48dB(A) in order to be acceptable for users [28], we believe our results reveal the ineffectiveness of a masking sound at the acceptable loudness levels in individual workspaces.

II. Background

A. Speech Masking Basics

Sound masking in the context of speech privacy refers to the process of hiding meaningful and sensitive human conversation from unwanted listeners. It can be useful in scenarios that require maintaining speech privacy, as well as in situations that demand a reduction of noise due to undesired human voices affecting productivity at the workplace.

Most privacy preserving mechanisms deployed in open office settings consist of physical partitions between workspaces of individual workers. While these methods do well at blocking unwelcome *visual* contacts, they are not very good at blocking *sounds*. In the open office setting, these partitions do not extend up to the ceiling or completely surround a workspace, allowing the sound waves to travel unimpeded along many directions throughout the workplace. In contrast, a personal office provides better overall privacy to an individual. However, it can not fully prevent sound from traveling outside or seeping into the room unless the closed office space is also sealed acoustically.

Both of the office scenarios described above show a need for an auxiliary privacy preserving mechanism in place that is able to prevent, to the most extent, the detrimental effect of speech movement across an individual's confidential space.

Most of the current sound masking solutions utilize a steady stream of ambient sound for hiding speech and providing a distraction free environment by reducing the speech to noise ratio, thereby reducing intelligibility [18], [24]. Increasing background noise level has the same effect on speech comprehension as does attenuating the speech signal [8]. Pink noise, in particular, has been the widely used background noise in these systems [13] and is defined as a signal having equal amount of energy in each octave.

Speech Intelligibility: The extent to which the speech privacy of a person is compromised depends upon multiple factors that determine the amount of meaningful speech that the unintentional listener is able to comprehend.

1) *Speaker's intensity:* The loudness of spoken words determines how far the sound will travel and how much of it would be intelligible.
2) *Attenuation:* It refers to the dissipation of energy in a signal as it travels through the transferring medium. Human sound after leaving the vocal tract of the speaker, travels

through the air and other physical obstacles like doors, walls, partitions etc. until it reaches the listener's ear. This will cause the sound to be attenuated and hence it will have a lower volume than at the origin.

3) *Background noise:* If background noise intensity is high, it will become very hard for the listener to make out the speech. If there is little background noise, the listener will have no trouble in deciphering the speech if it is near hearing levels.

B. Related Work

Cavanaugh et al.[8] studied the relation between speech privacy and speech intelligibility. They proposed a speech privacy index that rated the satisfaction of users about their privacy based on speech intelligibility in the presence of continuous background noise in an office space. Their measurement index was simplified by Young [31] and was modified to be used in open office designs by Pirn [21]. Egan [11] extended this work for the closed office design. Jensen et al. [16] analyzed acoustic satisfaction in office environments in the context of noise and speech privacy. They reported that people in closed offices are more satisfied with acoustics than in cubicles (semi-closed offices). They also found out that people in open office spaces felt satisfied with their acoustics and speech privacy as they could visually spot any unintended listener.

A spectrum for acceptable masking sound was proposed by Beranek [1]. Veitch et al. [27] reported that an optimum spectrum for masking sound should follow a similar slope as that of speech close to Brown noise (resembles waterfall or heavy rainfall). The effects of several masking sounds such as instrumental music, vocal music, water sounds, air ventilation system and filtered pink noise were studied by Haapakangas et al. [12]. They found that water sound followed by pink noise provided the most acoustic satisfaction while vocal music performed the worst. Hongisto et al. [13] revisited the issue and reported that sounds having mid-low frequencies were most satisfactory to users. Balancing speech masking and usability of the masking sound, they recommended a psuedorandom pink noise filtered to a slope of -7dB per octave to be used in open office environments.

III. Threat Model and Possible Attack Scenarios

Speech privacy mechanisms involve generating a masking sound to drown out any unintended acoustic leakage from an individual's confidential space. We will now describe scenarios that resemble some real world settings that may require the need for an evaluation of the deployed speech privacy tool based on sound masking.

Open office scenario: As discussed in Section I, an open office space is described as a workspace shared by multiple individual workers where the partitions defining the perimeter for each individual do not reach to the ceiling. This setting allows sound to travel around the boundaries of the partitions that are meant to provide privacy for each person. Examples of open office workplace include call centers, research areas, hospital wards, reception areas etc.

This setting introduces multiple situations where the individual may have their privacy compromised. As an example, the individual may be talking on the phone or having a private conversation with a coworker within the confinement of their space but their voices can be overheard by a malicious eavesdropper whose workspace is adjacent to the concerned workspace. Any person walking or standing in the vicinity of the speaker's workspace is also a potential eavesdropper. The perimeter of possible eavesdroppers in *open office* design is influenced by the loudness of the speaker.

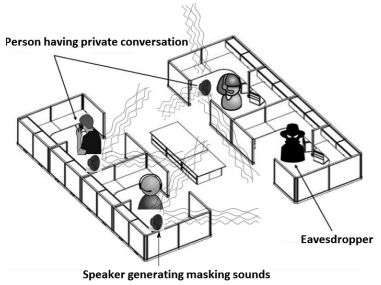
Private office scenario: In this scenario, the individual space is totally surrounded by walls and doors. There is little to no space between the door and the ground while the walls make sure that the acoustic leakage does not go unimpeded over or around them. While this setup is more privacy friendly than the *open office scenario* described previously, sound can still travel through walls, doors, and any gaps that exist between or around them.

In our threat model, we will not consider the materials used in building the walls or the door of the office. Similarly we will not discuss the acoustic design of the room or the workplace in the *private office scenario* and *open office scenario* respectively. The reason behind our exclusion of these factors is to isolate the vulnerability of the current sound masking approach. As stated in Section I, sound masking is used to hide unwanted sounds that have escaped even after application of "absorb" and "block" techniques. Hence, our experiments only consider the intelligibility of the leaked speech from the office space assuming "absorb" and "block" methods have already been implemented.

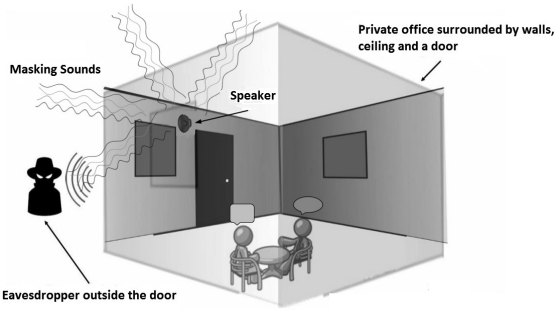
The eavesdropper in this scenario is a malicious entity whose intention is to learn sensitive information from the speaker's conversation. In our threat model, we place the eavesdropper in front of the door as its thickness is usually on par or less than the surrounding walls making it the most plausible leakage point. In addition, most of the offices are designed in fashion that places the occupant in front of the door making it directly in line with the direction in which the speaker's voice would be propagating.

In our work, we will only be analyzing masking solutions for individual workspaces. This class of sound masking solution provides the flexibility and portability to install the sound masking generator at a convenient location. These sound masking products for individual workspaces (Section I: *Our Contributions*) typically consist of a single masking sound generator system that can be turned on and off at an individual's preference. In our work, we focus only on sound masking and again, assume that other privacy measures such as "absorb" and "block" have already been implemented. This is an entirely plausible scenario as no measure can completely absorb or block sound except in acoustically sealed rooms.

An important component in our threat model is the equipment used for eavesdropping on speech. While many sophisticated devices are available that enable eavesdropping through doors and even walls, we restrict ourselves to generally available devices such as PC microphones (microphones used with



(a) Open office model



(b) Closed office model

Fig. 1: Threat model for different workplace scenarios

desktop computers) and smartphones for the eavesdropping device. The desktop microphones are a common commodity found in offices and are geared towards recording human voices. Most modern smartphones are equipped with fairly powerful microphones and their ubiquitous nature makes them a preferred tool for eavesdropping without raising any suspicion about the nefarious purpose of the eavesdropper.

Active noise cancellation: This technology is used for noise reduction in applications that necessitate high quality sound reception. The principle behind active noise cancellation involves estimating the existing background noise and generating an *anti-noise* signal similar to the existing background noise but inverted in phase. The *anti-noise* signal approximately cancels out the existing background noise to an extent, delivering a noise free signal. Since this technology already exists on smartphones that enjoy ease of access, we use it in our experiments to eavesdrop on leaked confidential speech.

Offline processing: In addition to *active noise cancellation* technology that can be used with an array of multiple microphones or the dual microphone setup found on smartphones, it also allows the eavesdropper to listen to the recorded speech samples multiple times offline. This advantage facilitates the eavesdropper to get familiar with recording and speculate on the spoken words. It also allows the eavesdropper to apply noise cancellation algorithms that use spectral subtraction among other techniques to remove noise from recorded samples and isolate vocals.

IV. Experiment Setup

In this section, we will draw out the scenarios where acoustic eavesdropping poses a security risk by compromising speech privacy. We will also discuss the masking signal used in our experiments and the motivation behind its usage. Finally, we will list other details about our experiment.

A. Tested Attack Scenarios

As discussed in Section III, a workplace design can roughly be categorized as open office space or private office space. In an open office space, the partitions between individual workspaces do not extend all the way up to the ceiling. On the other hand, a private office is sealed off from all sides with partitions (walls) that join the ceiling and a door, all of which are supposed to minimize audio and visual intrusion.

In our experiments, we designed three scenarios that represent various models for individual office space design. The first scenario is the open office scenario which offers the least obstruction of acoustic leakage from the workspace. For this setup, we chose an office space that lacks doors between adjacent rooms allowing sounds to travel unimpeded through the air across the rooms. This design offers the same advantage to an eavesdropper as the open office design where the partitions do not reach up to the ceiling, thereby allowing sounds to travel through the air gap between the partitions and the ceiling (e.g. cubicles). This setup is depicted in Figure 1a.

The second scenario is a private office design that we achieve by selecting an office room with walls on each side and a wooden door that reaches all the way to the floor with very little space between the floor and the door. This office space allows minimal sound leakage via the air and is the hardest to eavesdrop upon. This design is shown in Figure 1b.

The third and final scenario in our experiment is eavesdropping in a hybrid/semi-closed office room setting. This setup can be described as a mix of open office cubicle style and closed private office style. In this setup, the office space is surrounded by walls that join the ceiling similar to the closed private office setting but the door is not acoustically sealed in the manner that it does not reach all the way to the floor. This means that there exists a significant gap between the door and the floor through which sounds can travel unimpeded (directly or reflected) to the eavesdropper.

B. Masking Signal

Most of the current commercial speech masking systems have their patent sound masking algorithm but the desired property for a masking sound is to have a random neutral spectrum like pink noise as the masking signal in their speech privacy solutions [9]. White noise, on the contrary, is a poor choice for sound masking as it sounds “hissy” (like static from a radio) and does not provide effective coverage of the speech spectrum [18], [20]. Voicearrest Sound Masking system generates a masking sound similar to the whooshing sound of a high-end Heating, Ventillation and Air Conditioning (HVAC) system [26] that is spectrally similar to the masking sounds detailed above.

In order to closely resemble the noise generated by commercial solutions, we did an initial test with the AM1200, a self contained sound masking system from AtlasIED [14]. This device provides two types of masking signal, white noise and pink noise, which are played via inbuilt loudspeakers. In order to generalize the type of masking sound, we generated pink noise from Matlab using the Pink, Red, Blue and Violet Noise Generation tool [32]. We found that spectrum wise, both approaches produced similar pink noise hence we used the Matlab approach in our experiments.

C. Equipment

We played the pink noise through a portable Sony SRS-XB2 Bluetooth speaker while the speech samples were played by a Logitech z323 loudspeaker system that emulates live human speech. The loudspeakers allow us to modulate the loudness of the speech samples and its ability to reproduce low frequencies (bass effect) makes it a viable alternative to human speakers. The SRS-XB2 speaker has a frequency transmission range of 20 Hz – 20,000 Hz with a sampling rate of 44.1 kHz. The Logitech z323 system has a frequency response of 55 Hz – 20,000 Hz and is comprised of two speakers and a subwoofer. Sound pressure level was measured by an SLM305 digital sound level meter that has a frequency response of 31.5 Hz – 8500 kHz, measuring range of 30 dBA to 130 dBA, and a sampling rate of 2 times per second.

We tested four microphones as possible eavesdropping devices. Senal UB-440 is a USB condenser microphone with a frequency response of 50 Hz – 18,000 Hz with a sample rate of 48 kHz and Dynex USB microphone has a 150 Hz – 10,000 Hz frequency response. We also tested LG Nexus 5x and Samsung Galaxy S6 phones for eavesdropping on speech. Using a smartphone as an eavesdropping device provides the eavesdropper two advantages: a smartphone is a ubiquitous device and can be used covertly for eavesdropping, and smartphones have active noise cancellation in order to capture good quality speech. Both the LG Nexus 5X and Samsung Galaxy S6 have a dual microphone setup that is used for active noise cancellation as detailed in Section III. After initial tests in each scenario using both cellphones as the eavesdropping device, we found that there were no significant differences in the recording quality of each cellphone and therefore chose the Samsung Galaxy S6 as the eavesdropping device for the rest of our experiments.

D. Sound Pressure Level

Normal conversation levels in term of Sound Pressure Level (SPL) are estimated to be 40-60dB at a distance of 1 meter and ambient office environment to be 20-30dB [10], [23]. In our experiments, we maintained the distance of the speech source and the eavesdropper to be at 2 meters and the locations were devoid of any speech and other noise except the air conditioning vents with the SPL measurement to be 50dB. We measured SPL both at the eavesdropping location and and the speech source. At the speech source, it was set to be 70dB and at the eavesdropping location to 55dB representing a normal conversation at eavesdropping location. An SPL of 80dB at the speech source and 65dB at the eavesdropping location was denoted a loud conversation.

For the masking sound generator, we set the noise levels to be 55dB, 65dB and 75dB that represent scenarios where the masking sound is equal to or more than normal and loud conversations in the room. Since the ambient noise in our experiment locations was around 50dB, we also tested the eavesdropping in absence of any masking sound generator, relying upon the ambient noise to hide the speech. This testing also represents the case when the masking sound is softer/quieter than the conversation.

E. Speech Database

We used a phonetically balanced, US English single speaker CMU_Arctic speech dataset [3] consisting of both male and female speakers. The speech samples were around 5 seconds long containing an average of 10 words. For each noise level described in Section IV-D, we used four speech samples; two from the US_English_bdl (male) and two from the US_English_slt (female) databases. We also used different sets of voice samples for each location described in Section IV-A.

F. Noise Reduction

To counteract sound masking, deployed to prevent any unintended listeners from eavesdropping on confidential speech, we equip the eavesdropper with a noise cancellation ability. To have a more powerful solution for noise cancellation, we use the spectrum subtraction mechanism to reduce any consistent noise such as pink noise. VOICEBOX [7] is a speech processing toolbox implemented in Matlab that provides audio processing routines. We were particularly interested in speech enhancement routines: “specsub”, “spendered”, “ssubmmse” and “ssubmmsev”. We applied these routines on speech samples that were recorded under different noise levels (Section IV-D) and scenarios (Section IV-A). We performed an initial study to determine the most effective speech enhancement routine to be used in our next stages of the experiment. The factors determining effectiveness of the speech enhancement routine were intelligibility of post-processed speech sample and amount of artifacts [30] produced in the post-processed sample. We observed that “ssubmmse” performed the best out of the four routines described above based on our criteria and hence we chose to use it for filtering out noise in our experiment.

V. Speech Privacy against Human Listeners

Based on the threat model described in Section III and the experiment setup detailed in Section IV, we test the feasibility of speech eavesdropping in various workplace scenarios in the presence of masking noise. We chose graduate student lab space (office space with walls but lacking any doors between individual spaces) in the university for the *open office* scenario because it offers a very similar design. Additionally, we chose a private, restricted access lab room for the *closed office* scenario and a semi-private conference room for the *hybrid office* scenario.

A. Study Design

Our aim is to measure the intelligibility of eavesdropped speech, using human listeners, when the speech is recorded under different noise conditions as described in Section IV. We implemented the experiment in two phases: a recording phase and a comprehension phase. In the recording phase, we placed the eavesdropper at a distance of approximately 2 meters from a loudspeaker (Section IV-C) located centrally in the workspace. In the *open office* scenario, this placed the speaker in the middle of the room while the eavesdropper was outside in the adjacent workspace. For the *closed office* scenario, the speaker was placed inside the room centrally behind a closed door while the eavesdropper stood outside the door In the *semi-closed office* scenario, the speaker and the eavesdropper were situated similar to the *closed office* scenario.

For all three scenarios, the eavesdropper attempted to record speech with a Samsung Galaxy S6's IV-C microphone using the Smart Voice Recorder app with a sampling rate of 44.1 kHz. For the recording phase, two samples of male speech and two samples of female speech were played from the database (Section IV-E). After recording, we performed post-processing on the samples using Voicebox (Section IV-F). We listened to each post-processed sample and determined if any part of speech could be detected (not necessarily comprehended) in it. Samples containing only noise with no trace of human speech were marked as unusable.

For the comprehension phase, we used Amazon Mechanical Turk workers to listen to the recorded samples from the recording phase and write down the parts of speech that they could comprehend. We designed an online survey approved by our university's IRB that consisted of three stages, one for each of the three scenarios described in Section IV-A. We presented users of the survey with noisy speech samples recorded by the eavesdropper followed by the same set of noisy samples after applying noise reduction as per Section IV-F. This enabled us to determine the effect of noise reduction on the comprehension of noisy speech samples which in turn would indicate the effectiveness of masking sounds in preserving speech privacy.

Accuracy: We calculate accuracy for each sample as the ratio of number of words correctly inferred by listener to total number of words present in the speech sample.

B. Initial Observations from Post-processing

After post-processing with noise reduction, we marked each sample as usable or unusable based on traces of human speech contained in the sample. Our observations for each scenario are described below.

Open office scenario: We observed that when the noise level was 75dB, male and female speech samples recorded at 55dB at the eavesdropping site were unusable even after the post-processing phase. At a noise level of 65dB, male speech samples recorded at 55dB were unusable, but all samples were usable at a noise level of 55dB.

TABLE I: Average (and Maximum) accuracy for decoding words (0 represents lowest accuracy and 1 represents highest accuracy)

Speaker's loudness →	Male (65dB)	Female (65dB)	Male (55dB)	Female (55dB)
Noise level (75dB)	0.12 (0.62)	0.03 (0.23)	0.00 (0.00)	0.00 (0.00)
Noise level (65dB)	***0.40*** (0.92)	0.17 (0.62)	0.00 (0.00)	0.07 (0.44)
Noise level (55dB)	***0.90*** (1.00)	***0.51*** (1.00)	0.20 (0.69)	0.19 (0.54)
Noise level (50dB)	***0.86*** (1.00)	***0.83*** (1.00)	***0.82*** (1.00)	***0.55*** (1.00)

Closed office scenario: At a noise level of 75dB, male and female speech samples at 65dB and 55dB were deemed to be unusable. At a noise level of 65dB, male and female speech samples recorded at 55dB were also marked as unusable. However, at a noise level of 55dB, all samples were usable.

Hybrid office scenario: In this case, a noise level of 75dB made both male and female speech samples at 65dB and 55dB unusable. However, the other noise levels of 65dB and 55dB did not have any effect on the usability of the speech samples.

C. Results of Amazon Turk Survey

We collated the results from the amazon turk study and analyzed the users' response from various perspectives. The results contained responses from 29 human subjects (male 13; female 15; undisclosed 1) and are detailed for each of the scenarios below.

1) Open Office Scenario:: The average user response accuracy for the *open office* scenario are displayed in Table I. The results indicate that when the masking sound has a high loudness level (i.e. 75dB), users have difficulty comprehending any words. The maximum accuracy were only 0.12 or 12% when the spoken voice was from a loud male at 65dB. At a noise level comparable to conversation loudness at 65dB, the maximum accuracy for a male speaker at a similar loudness level jumped up to 0.40 or 40%. However female speaker's speech samples at the same loudness level were undecipherable. The male and female voice levels that are at normal loudness (55dB) could not be deciphered. This is to be expected as the masking sound would be louder than the speech.

When the masking sound level is decreased further to 55dB, the male speech samples at 65dB became almost completely comprehensible at 0.9 or 90% and the female voice samples at similar loudness also witnessed a jump in decoding accuracy up to 0.51 or 51%. The male and female speech samples that had the same loudness level as the masking sound were still not decipherable. Comparing these results to eavesdropping in the absence of any active masking sound generator (denoted by the last row in Table I), we observed that a male voice at the loud level was unaffected until the loudness of the masking sound was equal to or more than the male speaker's loudness. Female speaker's intelligibility was lower than male speaker's and was affected by masking sounds that were 10dB lower than their loudness level.

Table II denotes the accuracies after applying noise reduction as per Section IV-F. We excluded samples that were recorded in the absence of noise to restrict the effect of noise reduction when filtering the masking sound only. The results showed a minor increase in accuracies across different speech

TABLE II: Average (and Maximum) decoding accuracy after noise reduction

Speaker's Loudness→	Male (65dB)	Female (65dB)	Male (55dB)	Female (55dB)
Noise level (75dB)	0.14 (0.62)	0.03 (0.38)	0.00 (0.00)	0.00 (0.00)
Noise level (65dB)	***0.42*** (0.92)	0.20 (0.69)	0.00 (0.00)	0.12 (0.44)
Noise level (55dB)	***0.91*** (1.00)	***0.52*** (1.00)	0.25 (0.77)	0.22 (0.44)

TABLE III: Average (and Maximum) accuracy for decoding words (0 represents lowest accuracy and 1 represents highest accuracy)

Speaker's Loudness→	Male (65dB)	Female (65dB)	Male (55dB)
Noise level (65dB)	0.01 (0.18)	0.01 (0.14)	0.00 (0.00)
Noise level (55dB)	0.30 (0.91)	0.01 (0.86)	0.32 (0.17)
Noise level (50dB)	***0.73*** (1.00)	***0.95*** (1.00)	0.26 (0.92)

TABLE IV: Average (and Maximum) decoding accuracy after noise reduction (underlined numbers represent increase from Table III)

Speaker's Loudness→	Male (65dB)	Female (65dB)	Male (55dB)
Noise level (65dB)	0.03 (0.09)	0.05 (0.28)	0.00 (0.00)
Noise level (55dB)	***0.43*** (0.91)	0.01 (1.00)	***0.58*** (0.17)
Noise level (50dB)	***0.73*** (1.00)	***0.95*** (1.00)	0.26 (0.92)

TABLE V: Average (and Maximum) accuracy for decoding words (0 represents lowest accuracy and 1 represents highest accuracy)

Speaker's Loudness→	Male (65dB)	Female (65dB)	Male (55dB)	Female (55dB)
Noise level (65dB)	0.04 (0.50)	0.04 (0.50)	0.00 (0.00)	0.00 (0.00)
Noise level (55dB)	***0.58*** (1.00)	***0.59*** (1.00)	0.10 (0.67)	0.03 (0.33)
Noise level (50dB)	***0.82*** (1.00)	***0.93*** (1.00)	***0.91*** (1.00)	***0.63*** (1.00)

levels but did not show any dramatic improvements over the results obtained in Table I.

We also observed the maximum accuracy that could be obtained by decoding words from the speech samples under the influence of a masking sound. The results are shown within parenthesis in Table I. The figure indicates that the maximum accuracy even at a high loudness of masking sound (75dB) was over 60% for male speaker at 65dB and around 20% for female speaker at 65dB. The accuracies climbed higher when the masking sound level was lowered to 65dB with male speaker (65dB) reporting a maximum accuracy of over 90% and the female speaker (65dB) at around 60%. Even female speaker's speech samples at 55dB were decoded with a maximum accuracy of around 40%. Maximum accuracy after applying the noise reduction process detailed in Section IV-F is shown within parenthesis in Table II. This result is similar to the one shown in Table I so we see no major difference after noise reduction except in one case. There was an increase in the maximum accuracy for female speakers (65dB) in the presence of a masking sound at 75dB from 20% to 40%. A minor increase in accuracy exists for male speakers (55dB) in the presence of a masking sound at 55dB where it goes from around 70% to 77%.

2) Closed Office Scenario:: The results from the *closed office* scenario are shown in Table III. It indicates that the average accuracies are very low compared to the *open office* scenario. In the presence of a masking noise, the best accuracy was 30% when the masking sound level was 55dB and the speaker was male, speaking at 65dB. We did not consider female speakers at 55dB because it was too low to be heard in the presence of any level of masking sound at the eavesdropper's location. Also, loud masking sound at 75dB made any speech inaudible and hence was not included in our experiment with human subjects.

After noise reduction, we observe in Table IV that there is an increase in average accuracies with the best accuracy (58%) now that of the male speaker (55dB) in the presence of a masking noise at 55dB. There is also an increase in the decoding accuracy for male speakers at 65dB from 30% to 43% at masking sound (55dB).

The numbers within parenthesis in Table III show the maximum accuracy that we achieved among the human listeners that we used for decoding the speech samples in the *closed office* scenario. At a masking noise level of 65dB, we see that the best accuracy was below 20%. A male speaker at 55dB was not audible at this noise level and a female speaker was not audible at all noise levels. However, at a noise level of 55dB, we could decode almost 90% of spoken words for speakers louder than the masking sound i.e. at 65dB. For the male speaker at the same loudness as the masking sound (55dB), the accuracy was low at less than 20%. In the absence of any active masking sound generation, the speech was audible for all speakers with almost all words audible to them. When noise reduction is applied to the recorded speech samples in the *closed office* scenario as per Section IV-F, we see an increase in the decoding accuracies for the female speaker at both masking sound levels. For a masking sound level of 65dB, it increases to around 30% from 15% and at a masking sound level of 55dB, the increase is from 86% to 100%.

3) Hybrid/Semi-closed Office Scenario:: A *semi-closed office* scenario represents a setup between the two common workplace designs of *open office* and *closed office* scenarios. Table V suggests at a masking sound level of 55dB, almost half of the spoken words from male and female speakers in a loud voice (65dB) were successfully decoded. The accuracies were low for speakers at the same loudness as the masking sound (55dB). Similar to the *closed office* scenario, loud masking sound at 75dB outside the door completely obfuscated the speech at the eavesdropper location and hence was excluded from the comprehensibility test. In Table VI, we show the mean accuracies after applying noise reduction as detailed in Section IV-F. We observe an increase in accuracies for speakers at 65dB and for male speakers at 55dB. For speakers at 65dB in the presence of a low masking sound of 55dB, the mean accuracies now reach until 70% from 60%. Similarly we observe an approximately 10% increase in accuracies for male speakers at 55dB (masking sound at 55dB) and for speakers at 65dB (masking sound at 65dB).

The maximum accuracies for the *semi-closed office* scenario are shown within parenthesis in Table V. In the presence of a masking sound at 65dB for both male and female speakers at 65dB, 50% of the spoken words were decoded successfully. At a lower level of 55dB of masking sound, speech samples for loud speakers (65dB) were fully decipherable while male speakers at 55dB had a decoding accuracy of around 70% and female speakers at 55dB had an accuracy of approximately

TABLE VI: Average (and Maximum) decoding accuracy after noise reduction (underlined numbers represent increase from Table V)

Speaker's Loudness→	Male (65dB)	Female (65dB)	Male (55dB)	Female (55dB)
Noise level (65dB)	0.12 (0.50)	0.16 (0.83)	0.00 (0.00)	0.00 (0.00)
Noise level (55dB)	***0.69*** (1.00)	***0.73*** (1.00)	0.22 (0.89)	0.03 (0.33)
Noise level (50dB)	***0.82*** (1.00)	***0.93*** (1.00)	***0.91*** (1.00)	***0.63*** (1.00)

30%. After application of the noise reduction technique, there was a sizeable increase in the accuracy of female speakers at 65dB, in the presence of a masking sound at the same loudness, from 50% to over 80%. For a male speaker at 55dB in the presence of a masking sound at 55dB, we see an increase from approximately 70% to 90%.

VI. Speech Privacy against Automatic Speech Recognition

In this section, we will detail our experiments and results of using speech recognition tools to compromise speech privacy. We use the same threat model described in Section III outlining the attack scenarios for workplace environments.

A. Automatic Speech Recognition

An automatic speech recognition mechanism converts a recorded audio signal into words. In modern automatic speech recognition methodology, the aim is to infer spoken words from the given signal by using neural networks that utilize Hidden Markov Models (HMM) combined with acoustic models and based on Gaussian Mixture Models (GMM). A more recent trend in the industry has been to replace the GMM based models with deep neural network (DNN) learning [15], [19] that was first introduced in [29]. A common use of automatic speech recognition today is in recognizing voice commands and queries from a user in different environments. From a usability perspective, such systems should be able to recognize voice commands under varying scenarios that typically include background noise and different types of user's speech. An effective way of successfully performing speech recognition under adverse conditions is to use DNN and train it on a vast set of pre-collected noisy samples. We utilize this method for compromising speech privacy by feeding such speech recognition tools with audio samples that were eavesdropped under background noise produced by sound masking systems.

B. Experiment Setup

In our automatic speech recognition setup, we reuse the speech samples that were collected in Section V-A for the three scenarios: *open office*, *closed office* and *semi-closed office*. Both raw speech samples and speech samples post processed with Voicebox, to limit the effect of background noise on the comprehensibility of the spoken words, were used for determining the effectiveness of automatic speech recognition systems in this endeavor.

We tested the Google Cloud Speech Recognition service, Microsoft Bing Voice Recognition service, PocketSphinx and the IBM Watson Speech to Text service in our speech recognition task. Our results indicated that Google Cloud Speech Recognition service performed best while Microsoft Bing Voice Recognition, IBM Watson Speech to Text service and PocketSphinx performed poorly being unable to recognize the majority of the spoken words in our samples. Hence, we report the speech recognition results from using Google's services as they were the most accurate under noisy environment. Google Cloud Speech Recognition claims to perform accurately for noisy audio which also makes it a suitable candidate for our experiments.

TABLE VII: Average accuracy for decoding words (0 represents lowest accuracy and 1 represents highest accuracy)

Speaker's loudness →	Male (65dB)	Female (65dB)	Male (55dB)	Female (55dB)
Noise level (75dB)	0.00	0.00	0.00	0.00
Noise level (65dB)	0.00	0.31	0.00	0.11
Noise level (55dB)	0.92	0.61	0.00	0.00
Noise level (50dB)	0.92	0.92	0.92	0.33

C. Results from Automatic Speech Recognition

We report the results from the three scenarios for both noisy speech samples and speech samples generated after noise subtraction. We use a method similar to the one described in Section V-A to calculate the accuracy of a speech recognition system in transcribing words from the audio samples.

1) Open Office Scenario:: Our results for decoding words from the recorded audio sample in the *open office* scenario are tabulated in Table VII. It can be observed that the Google Cloud Speech Recognition service was able to recover almost full speech (92%) from eavesdropped sentences in the audio samples for all speakers (except normal female voice at 55dB that produced an accuracy of 33%) at both normal and loud acoustic levels in the absence of any artificial noise.

In the presence of an artificial noise generator producing static Gaussian noise as described in Section IV-C at 55dB, we noticed a decrease in the decoding accuracy that dropped to 0% for male and female voices at a similar loudness level as the generated noise. The loud male voice (65dB) did not get affected though there was a decrease in accuracy of loud female voice (65dB) dropping from 92% to 61%. Increasing the level of noise in the recorded audio samples indicated a further decrease in the decoding accuracy with the speech recognition tool almost completely failing when the noise level was very loud (75dB).

TABLE VIII: Average decoding accuracy after noise reduction

Speaker's loudness →	Male (65dB)	Female (65dB)	Male (55dB)	Female (55dB)
Noise level (75dB)	0.00	0.00	0.00	0.00
Noise level (65dB)	0.08	0.08	0.00	0.00
Noise level (55dB)	0.85	0.31	0.00	0.22

After applying noise reduction using Voicebox as explained in Section IV-F, we tested the processed audio samples against the speech recognition tool again. The results are shown in Table VIII and indicate that applying of noice reduction actually degraded the quality of human voice in the audio sample with a decrease in recognition accuracy for loud male voice from 92% to 85% and for loud female voice from 61% to 31% at a noise level of 55dB. This behavior may be due to artifacts produced in the audio sample during noise reduction

TABLE IX: Average accuracy for decoding words (0 represents lowest accuracy and 1 represents highest accuracy)

Speaker's loudness →	Male (65dB)	Female (65dB)	Male (55dB)	Female (55dB)
Noise level (75dB)	0.00	0.00	0.00	0.00
Noise level (65dB)	0.00	0.00	0.00	0.00
Noise level (55dB)	0.45	0.28	0.00	0.00
Noise level (50dB)	0.91	1.00	0.00	0.00

that uses spectrum subtraction. This procedure seemed to adversely affect the accuracy of the speech recognition system that tends to rely on its pre-trained dataset for deciphering speech which does not work well with artificial post processing side-effects.

2) Closed Office Scenario:: In the *closed office* scenario, the accuracies for decoding words from the noisy audio samples are shown in Table IX. The speech recognition system failed to decipher any words from the audio samples for normal human voices (both male and female). For loud voices at 65dB, male and female spoken words were almost completely decoded in an ambient noise atmosphere at 50dB. In the presence of a noise generating device, the decoding accuracies dropped to 45% from 92% for the male voice and 28% from 100% for the female voice at a noise level of 55dB. There was a complete failure of speech recognition at higher loudness levels of noise. For the processed audio samples (where noise reduction is performed), the speech recognition tool failed to detect any human voice in this scenario. Thus we determine that the speech recognition system can not be used effectively in conjunction with a noise reduction method like spectrum subtraction in the *closed office* scenario where the SNR is already low and applying noise reduction may harm the human voice segments in the recorded audio.

3) Hybrid/Semi-closed Office Scenario:: The results for the *semi-closed office* scenario are depicted in Table X and show that the Google Cloud Speech Recognition service was able to detect male voices at 65dB with 67% and at 55db with 55% accuracy. The accuracies were higher for female voices: 92% at 65dB and 100% at 55dB. In the presence of active noise generation at 55dB, these accuracies quickly dropped with complete recognition failure at the normal voice levels of 55dB. The accuracies at a loudness of 65dB were 17% for male voices and 58% for female voices. No words were detected at higher noise levels for any of our samples.

TABLE X: Average accuracy for decoding words (0 represents lowest accuracy and 1 represents highest accuracy)

Speaker's loudness →	Male (65dB)	Female (65dB)	Male (55dB)	Female (55dB)
Noise level (75dB)	0.00	0.00	0.00	0.00
Noise level (65dB)	0.00	0.00	0.00	0.00
Noise level (55dB)	0.17	**0.58**	0.00	0.00
Noise level (50dB)	0.67	0.92	0.55	1.00

VII. Discussion

Summary and Reflection on Results: We showed that sound masking using pink noise and other similar masking sounds may not be suitable for preserving speech privacy in workplace environments. Using human subjects as decoders, we showed that in an open office environment, almost 90% of speech could be deciphered when masking sound was 10dB lower than the speaker. We also observed that it was easier to decipher male speech in a noisy environment than female speech at the same loudness level and under similar noisy conditions. The average accuracy for decoding loud male voices was around 90% and for a female voice at the same loudness level and noise conditions, it was 50%. We attribute this result to the differences between male and female voices as male voices contain lower frequencies due to heavier and larger vocal folds and hence travel farther and suffer less attenuation.

We also measured the maximum accuracy that tends to exploit the best comprehension ability among all the human workers and represents the maximum degree of speech privacy threat. We found out that at a low level of sound masking at 55dB, loud human voices (65dB) could be completely deciphered while softer voices (55dB) could be decoded with a reasonable accuracy. At 65dB of masking sound, loud human voices could still be heard and comprehended denoting the fact that any masking sound used should be louder than speech. We also noted that in this setting, noise reduction tools only marginally increased the accuracy except in the case of female speakers (65dB) in the presence of a masking sound at 55dB where the maximum accuracy gets doubled from 20% to 40%.

Results from the *closed office* scenario show that it is more difficult to eavesdrop on speech when compared to the *open office* scenario. The average accuracies are around 30% or lower. Application of noise reduction techniques however increase the accuracies to 40% and above for male speakers. In this scenario, the female speaker's voice is hardly audible at 65dB and 55dB when compared to male speech. This observation is in line with the observations made in the *open office* scenario described previously. The maximum accuracies in this scenario indicate that the *closed office* scenario is harder to eavesdrop upon and a masking sound makes it harder for an eavesdropper to decode the speech. Still, if the speech is louder than the masking sound (a difference of 10dB in our experiments), it is possible to decode 90% of the speech. Using a noise reduction process, this accuracy can be increased to 100%.

A *semi-closed office* scenario is more vulnerable to eavesdropping attacks than a *closed office* scenario. The average accuracies indicate that almost half of the speech can be decoded if the speaker's voice is louder (+10dB in our experiments) than the masking sound. Noise reduction can improve this accuracy to 70%. The maximum accuracies indicate that speech privacy can be totally compromised in this scenario if the speaker's voice is louder than the masking sound (+10dB in our experiments). For speakers of comparable loudness, the accuracies are still high for male speakers at around 70% while for female speakers, the accuracy is around 30%. Noise reduction improves the accuracy for male speakers at 55dB to almost 90dB thereby increasing the threat level significantly in the case of comparable loudness level of the speaker and masking noise.

Our results from offline processing of the audio samples

using the Google Cloud Speech API indicate that such systems still may not be as effective at determining speech in noisy environments as human listeners. These systems completely failed to detect speech as the SNR approached 1, in all workplace scenarios. In our work, application of spectrum subtraction reduced noise and some portion of the speech spectrum which resulted in deterioration of the recognition accuracy. Thus these tools may perform reasonably well compared to human listeners, but their performance is worse after noise reduction.

Potential Countermeasures: In order to counteract attacks that exploit the weakness of the noise generating mechanism, careful steps need to be taken that comply with the privacy demands of the scenario. Proper efforts must be made to ensure that the SNR in the environment remains close to 1 and the system must maintain usability for the victim. Since we showed that pseudo-random noise is susceptible to filtering by spectrum subtraction and by active noise cancellation (used extensively in smartphones), it may be fruitful to revisit the frequency spectrum of masking sound and possibly redesign the masking signal with a dynamic frequency spectrum. In addition to masking sounds based on pink or brown noise, other options must be explored that may provide a more robust coverage against such eavesdropping attacks.

VIII. Conclusion

In this work, we showed that individual workplaces deploying commercial sound masking solutions which use psuedo-random noise for speech privacy may be vulnerable against an eavesdropping adversary. We observed that in different workplace settings for individual offices, such psuedo-random noise can be effectively counteracted by using low cost, off the shelf devices such as smartphones. Additionally, spectral subtraction algorithms can be used for noise reduction, further compromising the speech privacy. Based on our results, we believe that psuedo-random noise may not be suitable for speech masking and masking sounds need to be redesigned in a manner that is robust against the mentioned attacks.

References

[1] Beranek, L.L., Blazier, W.E.: Preferred noise criterion (pnc) curves and their application to rooms. Journal of the Acoustical Society of America **50(5)**, 1223–1228 (1971)

[2] Berger, T.: Speech privacy and sound masking in modern architecture (2015), https://www.bicsi.org/uploadedfiles/BICSI_Conferences/Canada/2015/presentations/Speech_Privacy.pdf

[3] Black, A.W.: Cmu_arctic speech synthesis databases (2017), http://festvox.org/cmu_arctic/

[4] Brand, J.L., Smith, T.J.: Effects of reducing enclosure on perceptions of occupancy quality, job satisfaction, and job performance in open-plan offices. In: Proceedings of the Human Factors and Ergonomics Society. pp. 818–822 (2005), www.scopus.com

[5] Brennan, A., Chugh, J.S., Kline, T.: Traditional versus open office design. Environment and Behavior **34**(3), 279–299 (2002)

[6] Brill, M., Weidemann, S., Associates, B.: Disaproving Widespread Myths about Workplace Design. Kimaball International, New York (2001)

[7] Brookes, M.: Voicebox: Speech processing toolbox for matlab (2017), http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html

[8] Cavanaugh, W., W.R., F., P.W., H., B.G., W.: Speech privacy in buildings. Journal of the Acoustical Society of America **34**, 475–492 (1962)

[9] Chanaud, R.: Sound Masking Done Right: Simple Solutions for Complex Problems. Magnum Publishing L.L.C. (2008), http://www.atlasied.com/f/AtlasSound/SoundMaskingDoneRight.pdf

[10] Department, E.P.: Characteristics of sound and the decibel scale (2017), http://www.epd.gov.hk/epd/noise_education/web/ENG_EPD_HTML/m1/intro_5.html

[11] Egan, M.D.: Concepts in Architectural Acoustics. McGraw Hill, New York, NY (1972)

[12] Haapakangas, A., Kankkunen, E., Hongisto, V., Virjonen, P., Oliva, D., Keskinen, E.: Effects of five speech masking sounds on performance and acoustic satisfaction. implications for open-plan offices. Acta Acustica united with Acustica **97**(4), 641–655 (2011), http://www.ingentaconnect.com/content/dav/aaua/2011/00000097/00000004/art00011

[13] Hongisto, V., Oliva, D., Rekola, L.: Subjective and objective rating of spectrally different pseudorandom noises—implications for speech masking design. The Journal of the Acoustical Society of America **137**(3), 1344–1355 (2015). https://doi.org/10.1121/1.4913273, http://dx.doi.org/10.1121/1.4913273

[14] IED, A.: Self contained sound masking system ul2043 with built in loudspeakers (2017), https://www.atlasied.com/low-profile-sound-masking-system-ul2043

[15] Jaitly, N., Nguyen, P., Senior, A., Vanhoucke, V.: Application of pretrained deep neural networks to large vocabulary speech recognition. In: Proceedings of Interspeech 2012 (2012)

[16] Jensen, K.and Arens, E.: Acoustical quality in office workstations, as assessed by occupant surveys. In: Proceedings of Indoor Air 2005. Beijing, China (2005)

[17] Kaarlela-Tuomaala, A., Helenius, R., Keskinen, E., Hongisto, V.: Effects of acoustic environment on work in private office rooms and open-plan offices – longitudinal study during relocation. Ergonomics **52**(11), 1423–1444 (2009)

[18] Management, C.S.: Sound masking 101 (2017), http://cambridgesound.com/learn/sound-masking-101/

[19] rahman Mohamed, A., Dahl, G.E., Hinton, G.: Acoustic modeling using deep belief networks. IEEE Transactions on Audio, Speech, and Language Processing **20**(1), 14–22 (2012)

[20] Network, L.A.: Sound masking (2017), https://www.logison.com/technology/sound-masking

[21] Pirn, R.: Acoustical variables in open planning. Journal of the Acoustical Society of America **49(5), part 1**, 1339–1345 (1971)

[22] Salter, C., Powell, K., Begault, D., Alvarado, R.: Case studies of a method for predicting speech privacy in the contemporary workplace. Tech. rep., Center for the Built Environment, UC Berkeley (2003)

[23] Sengpiel, E.: Decibel table–loudness comparison chart (2011), http://www.siue.edu/~gengel/ece476WebStuff/SPL.pdf

[24] Sound, A.: Sound masking achieving speech privacy cost effectively (2012), https://www.atlasied.com/f/2795/Atlas\%20Sound\%20Masking\%20Data\%20Sheet.pdf

[25] Systems, S.P.: How sound masking works (2017), https://www.speechprivacysystems.com/how-sound-masking-works/

[26] Systems, S.P.: Voicearrest sound masking system (2017), https://speechprivacysystems.com/voicearrest-sound-masking-system-2/

[27] Veitch, J., Bradley, J., Legault, L., Norcross, S., Svec, J.: Masking speech in open-plan offices with filtered pink noise noise: Noise level and spectral composition effects on acoustic satisfaction. Internal Report IRC-846, Institute for Research in Construction, Ottawa, Canada (2002)

[28] Veitch, J., Bradley, J., Legault, L., Norcross, S., Svec, J.: Masking speech in open-plan offices with simulated ventilation noise level and spectral composition effects on acoustical satisfaction (April 2002)

[29] Waibel, A., Hanazawa, T., Hinton, G., Shikano, K., Lang, K.J.: Readings in speech recognition pp. 393–404 (1990), http://dl.acm.org/citation.cfm?id=108235.108263

[30] Wikipedia: Sonic artifact (2017), https://en.wikipedia.org/wiki/Sonic_artifact

[31] Young, R.W.: Re-vision of the speech-privacy calculation. Journal of the Acoustical Society of America **38**, 524–530 (1965)

[32] Zhivomirov, H.: Pink, red, blue and violet noise generation with matlab implementation version 1.5 (2017), http://www.mathworks.com/matlabcentral/fileexchange/42919-pink--red--blue-and-violet-noise-generation-with-matlab-implementation?focused=6636118\&tab=function