MULTIPATH-ENABLED PRIVATE AUDIO WITH NOISE

Anadi Chaman, Yu-Jeh Liu*, Jonah Casebeer†, Ivan Dokmanić**

Departments of *Electrical and Computer Engineering and †Computer Science
University of Illinois at Urbana-Champaign

ABSTRACT

We address the problem of privately communicating audio messages to multiple listeners in a reverberant room using a set of loudspeakers. We propose two methods based on emitting noise. In the first method, the loudspeakers emit noise signals that are appropriately filtered so that after echoing along multiple paths in the room, they sum up and descramble to yield distinct meaningful audio messages only at specific *focusing spots*, while being incoherent everywhere else. In the second method, adapted from wireless communications, we project noise signals onto the nullspace of the MIMO channel matrix between the loudspeakers and listeners. Loudspeakers reproduce a sum of the projected noise signals and intended messages. Again because of echoes, the MIMO nullspace changes across different locations in the room. Thus, the listeners at focusing spots hear intended messages, while the acoustic channel of an eavesdropper at any other location is jammed. We show, using both numerical and real experiments, that with a small number of speakers and a few impulse response measurements, audio messages can indeed be communicated to a set of listeners while ensuring negligible intelligibility elsewhere.

***Index Terms*—** Private audio communication, speech privacy, multi-channel convolutional synthesis, speech intelligibility.

1. INTRODUCTION

Consider the problem of sending audio messages to different listeners in a reverberant room, while making sure that each message can only be understood by its intended recipient. Importantly, no eavesdropper anywhere in the room should be able to understand any of the messages.

This problem is related to personal audio zones and sound field reproduction [1–9] where the goal is to reproduce different sound streams in a few predefined *zones* in a room while minimizing the sound level everywhere else. In most of these approaches, however, an eavesdropper with a sensitive microphone (or a good ear) can easily understand the messages. The reason is that the loudspeakers simply reproduce linearly filtered versions of desired messages which remain highly correlated with any residual error signal.

To address the problem of private audio communication, we propose two methods. As an extension of our previous work [10], the first approach communicates audio messages to intended *focusing spots* by emitting appropriately filtered white Gaussian noise signals from loudspeakers. The filters are constructed such that after passing through specific sets of paths and time delays, these filtered random signals sum up coherently as they arrive at the target focusing points. On the other hand, they yield incoherent signals at locations with different sets of signal propagation paths. This solution is expected to work well when a room has high spatial diversity of acoustic channels.

In our second approach, the idea is to send random noise from loudspeakers in addition to message signals, such that the noise signals add up to zero only at the intended listening points, while they continue to mask the messages everywhere else. This results in the interception of clean audio messages at the focusing spots while having low intelligibility at other locations. This technique is inspired by standard methods in wireless networking on jamming eavesdroppers [11,12]. However, to the best of our knowledge, the prior works consider fading wireless channels without explicitly considering intersymbol interference (echoes). While this could be a fair assumption for networks like WiFi where sampling times are much larger than propagation delays of wireless signals, this is not the case in room acoustics. Hence, we adapt this jamming scheme to work with long convolutional channels.

Privacy in multizone reproduction systems was first studied in [13] where the authors also use noise to mask message signals in "quiet" zones to reduce intelligibility. While their method is applicable in both anechoic and reverberant conditions, the performance is degraded in the presence of echoes. On the other hand, as we elaborate later, our methods critically rely on echoes and multipath propagation. In particular, our solutions exploit the spatial diversity of room impulse responses (RIRs) across different locations in a room and the redundant degrees of freedom in signal transmission provided by multiple loudspeakers. Unlike in multizone methods, however, we can only deliver messages to a small, fixed region of space. On the other hand, we achieve good performance using a rather small number of loudspeakers and impulse response measurements (in our experiments we use only six).

The problem of jamming eavesdroppers has been studied extensively in wireless communication. The theoretical foundation was laid by Shannon [14] and later extended by [15, 16] who showed the feasibility of secrecy if the communication channel of an eavesdropper is degraded. The methods in [11, 12, 17] use artificial noise; [18] showed the possibility of secret communication as a consequence of slow wireless fading. Prior works have also looked at a related problem of eavesdropper detection [19–21].

In this paper, we empirically show that unlike traditional multizone sound field reproduction which is usually degraded in reverberant environments [22, 23], both of our proposed approaches give excellent results in the presence of echoes since echoes enhance spatial diversity. We derive conditions needed to generate desired messages at the focusing spots, and demonstrate both numerically and through real experiments that with six speakers and the knowledge of RIRs at the intended listening points, private audio communication is effectively achievable. In addition, we compare the robustness of the two approaches to system failures and uncertainties.

Project webpage: https://swing-research.github.io/private-audio/

2. PROBLEM FORMULATION

Consider a system with L loudspeakers, each emitting an audio signal to K listeners. Without loss of generality, let the desired length of the signal $\boldsymbol{y}_k$ at the k^{th} listener be N. We also assume that the room impulse response (RIR) between the k^{th} listener and the i^{th} speaker is a sequence h_{ki} which is L_{h} long and known a priori.

This signal received by the k^{th} listener is given as a sum of convolutions:

$$y_k(n) = \sum_{i=1}^{L} (h_{ki} * x_i)(n),\ n = 0, 1, ..., N-1, \tag{1}$$

where $\boldsymbol{x}_i \in \mathbb{R}^{L_{\text{x}}}$ is the signal transmitted by the i^{th} speaker with length $L_{\text{x}} = N - L_{\text{h}} + 1$, and $*$ represents linear convolution. We define intended message vector $\boldsymbol{y}_{\text{in}} \in \mathbb{R}^{NK}$ as a concatenation of all $\boldsymbol{y}_k \in \mathbb{R}^N$: $\boldsymbol{y}_{\text{in}} = [\boldsymbol{y}_1^\top, \boldsymbol{y}_2^\top, \dots \boldsymbol{y}_K^\top]^\top$. Similarly, we define channel matrices $\boldsymbol{H}_k$ of size $N \times LL_{\text{x}}$ as $[\boldsymbol{H}_{k1}, \boldsymbol{H}_{k2}, \dots, \boldsymbol{H}_{kL}]$, where each $\boldsymbol{H}_{ki}$ is a Toeplitz convolution matrix composed using $\boldsymbol{h}_{ki}$. Defining $\boldsymbol{H} = [\boldsymbol{H}_1^\top, \boldsymbol{H}_2^\top, \dots, \boldsymbol{H}_K^\top]^\top$ and $\boldsymbol{x} = [\boldsymbol{x}_1^\top, \boldsymbol{x}_2^\top, \dots, \boldsymbol{x}_L^\top]^\top$, (1) can be rewritten as:

$$\boldsymbol{y}_{\text{in}} = \boldsymbol{H}\boldsymbol{x}. \tag{2}$$

If the matrix $\boldsymbol{H}$ has full row rank, we can reconstruct any desired message signals at the K listeners. A well-known solution to (2) is given by $\boldsymbol{x} = \boldsymbol{H}^\dagger \boldsymbol{y}_{\text{in}}$, where $\boldsymbol{H}^\dagger$ is the pseudoinverse of $\boldsymbol{H}$. Though this solution suffices for message reconstruction at the listeners, it does not enforce unintelligibility at other locations. We could, however, exploit the additional degrees of freedom provided by the nullspace of $\boldsymbol{H}$ to generate a suitable $\boldsymbol{x}$ that ensures signal degradation outside the target focusing spots.

We note that for typical audio sampling rates, RIR lengths and message lengths, $\boldsymbol{H}$ is far too large to compute the pseudoinverse explicitly. That is why we solve all least-squares design problems in this paper by the conjugate gradient method. Since the involved matrices are all block-Toeplitz, the conjugate gradient method can be efficiently implemented using fast Fourier transforms.

3. THE TWO APPROACHES

As per (2), $\boldsymbol{x}$ can be suitably chosen to ensure that the message signals outside the focusing spots remain unintelligible. In this section, we present two methods to achieve this task, each constructing $\boldsymbol{x}$ in a different way: (i) multichannel convolutional synthesis (MCCS) by noise and (ii) noise in the nullspace approach.

3.1. Multichannel convolutional synthesis by noise

Recall from (1) that the signal arriving at the k^{th} listener is $\boldsymbol{y}_k = \sum_{i=1}^{L} \boldsymbol{h}_{ki} * \boldsymbol{x}_i$. In this first approach, we constrain $\boldsymbol{x}_i$ to be a convolution of a filter $\boldsymbol{g}_i$ of length L_{g} with a noise signal $\boldsymbol{n}_i$ of length L_{n}, drawn from standard normal distribution. This is equivalent to

$$\boldsymbol{x}_i = \boldsymbol{N}_i \boldsymbol{g}_i,\ i = 1, 2, \dots, L, \tag{3}$$

where $\boldsymbol{N}_i$ is an $L_{\text{x}} \times L_{\text{g}}$ Toeplitz convolution matrix composed using the vector $\boldsymbol{n}_i$, with $L_{\text{x}} = L_{\text{g}} + L_{\text{n}} - 1$. We define $\boldsymbol{g} = [\boldsymbol{g}_1^\top, \boldsymbol{g}_2^\top, \dots, \boldsymbol{g}_L^\top]^\top$ and a block diagonal matrix $\boldsymbol{N}$ as

$$\boldsymbol{N} = \text{diag}([\boldsymbol{N}_1, \boldsymbol{N}_2, \dots, \boldsymbol{N}_L]).$$

Then equations in (3) can be combined for all $i \in \{1, \dots, L\}$ to give $\boldsymbol{x} = \boldsymbol{N}\boldsymbol{g}$ and

$$\boldsymbol{y}_{\text{in}} = \boldsymbol{H}\boldsymbol{N}\boldsymbol{g}. \tag{4}$$

Given $\boldsymbol{H}\boldsymbol{N}$ and $\boldsymbol{y}_{\text{in}}$, $\boldsymbol{g}$ can be computed using conjugate gradient method.

This model constrains $\boldsymbol{x}$ to lie on a subspace of random vectors. To understand why, consider the signal emitted by the i^{th} loudspeaker, $\boldsymbol{x}_i$, which can be written as

$$x_i(n) = \sum_{p=0}^{L_{\text{n}}-1} n_i(p) g_i(n-p),\ n = 0, 1, ..., L_{\text{x}} - 1.$$

We can interpret $\boldsymbol{x}_i$ as a sum of randomly-scaled translates of filter $\boldsymbol{g}_i$. For all speakers, $\boldsymbol{g}_i$ are constructed such that convolutions of $\boldsymbol{x}_i$ with room impulse responses sum up to yield the desired messages only at the listeners. Thus, a specific set of RIRs $\{\boldsymbol{h}_{ki}\}$, corresponding to the intended listener–speaker pairs correctly descrambles the translates. In a room with rich spatial diversity, locations other than the intended listening points will be characterized by a different set of RIRs. We thus cannot expect the descrambling to yield the correct output, and the randomness of $\boldsymbol{n}_i$ then ensures non-intelligibility of the resulting signal.

3.2. Noise in the nullspace

We adapt the second approach from the wireless communications literature. Concretely, $\boldsymbol{x}$ is chosen as a sum of a message-carrying vector $\boldsymbol{s} \in \mathbb{R}^{LL_{\text{x}}}$ and a noise-like signal $\boldsymbol{w} \in \mathbb{R}^{LL_{\text{x}}}$, i.e., $\boldsymbol{x} = \boldsymbol{s} + \boldsymbol{w}$. We construct $\boldsymbol{s}$ and $\boldsymbol{w}$ to satisfy $\boldsymbol{H}\boldsymbol{s} = \boldsymbol{y}_{\text{in}}$ and $\boldsymbol{H}\boldsymbol{w} = \boldsymbol{0}$, so that

$$\boldsymbol{y}_{\text{in}} = \boldsymbol{H}(\boldsymbol{s} + \boldsymbol{w}) = \boldsymbol{H}\boldsymbol{s}. \tag{5}$$

This is achieved by choosing $\boldsymbol{w}$ as the projection of a random noise vector on the nullspace of the channel matrix $\boldsymbol{H}$, i.e., $\boldsymbol{w} = \boldsymbol{P}_{\mathcal{N}(\boldsymbol{H})}\boldsymbol{v}$, where the entries of $\boldsymbol{v}$ are i.i.d. standard Gaussian and $\boldsymbol{P}_{\mathcal{N}(\boldsymbol{H})}$ is the projector on the null space of $\boldsymbol{H}$.

As mentioned in Section 2, $\boldsymbol{H}$ is typically large, which makes the direct computation of its nullspace a prohibitively complex task. Instead, we first find the projection of $\boldsymbol{v}$ on the row space of $\boldsymbol{H}$ by solving

$$\hat{\boldsymbol{z}} = \underset{\boldsymbol{z}}{\text{argmin}} \|\boldsymbol{v} - \boldsymbol{H}^\top \boldsymbol{z}\|_2^2. \tag{6}$$

We again use the conjugate gradient method to solve (6) using fast Fourier transforms since $\boldsymbol{H}$ is block-Toeplitz. Once $\hat{\boldsymbol{z}}$ is found, the nullspace projection $\boldsymbol{P}_{\mathcal{N}(\boldsymbol{H})}\boldsymbol{v}$ is simply $\boldsymbol{v} - \boldsymbol{H}^\top \hat{\boldsymbol{z}}$.

4. CONDITIONS FOR PERFECT RECONSTRUCTION

In this section, we present the conditions needed to ensure perfect reconstruction of any set of message signals of length N at the K listeners (or any $\boldsymbol{y}_{\text{in}} \in \mathbb{R}^{NK}$) for both approaches.

4.1. Multi-channel convolutional synthesis by noise

From (4), perfect reconstruction can be achieved if the overall channel matrix $\boldsymbol{H}\boldsymbol{N}$ has full row rank, NK.

We make the assumption that the room is drawn randomly from a continuous distribution. (For example, let the corners be chosen uniformly at random within fixed balls.) We also assume that the loudspeaker and listener positions are placed at random according to an absolutely continuous distribution. These assumptions imply that the distribution of the nullspace of $\boldsymbol{H}$ is absolutely continuous with respect to the Haar measure on the Grassmannian. Then, we have the following result.

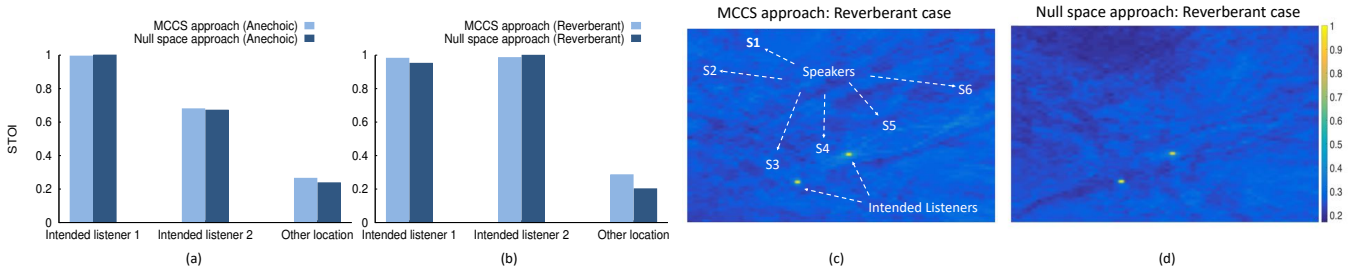

Fig. 1: STOI scores at 2 intended listeners and one additional location using MCCS and nullspace approach in (a) anechoic and (b) reverberant setting. (c)-(d) Heat maps reflecting STOI scores at 4200 locations in a simulated room of size 7 m×8 m. Speakers illustrated as S1-S6.

Proposition 4.1. *Suppose* $LL_g \geq NK$. *Then* $\boldsymbol{HN}$ *has full row rank with probability* 1.

Proof. We have that $\text{rank}(\boldsymbol{HN}) \leq \min\{\text{rank}(\boldsymbol{H}), \text{rank}(\boldsymbol{N})\}$ by rank inequalities. With the conditions of the proposition, this implies that $\text{rank}(\boldsymbol{HN}) \leq NK$. The only way to have a strict inequality is that the nullspace of $\boldsymbol{H}$ intersects the range of $\boldsymbol{N}$ along a subspace of dimension greater that $LL_g - NK$. On the other hand, because the nullspace of $\boldsymbol{H}$ is continuously distributed and independent from $\boldsymbol{N}$, it will intersect the range of $\boldsymbol{N}$ exactly along a subspace of dimension $LL_g - NK$ with probability 1. □

This result implies that for most setups in sufficiently reverberant rooms, we will be able to produce the desired messages at the listener positions.

4.2. Noise in nullspace approach

From (5), $\boldsymbol{H}$ needs to have full row rank for perfect reconstruction of all $\boldsymbol{y}_{\text{in}} \in \mathbb{R}^{NK}$. Similar to the previous case, since $\boldsymbol{H}$ is a function of the RIRs between the speaker-listener pairs, it is not completely in the user's control to ensure that it has full rank as it depends on room geometry and the spatial diversity of RIRs. In practice, however, if we assume a randomized setup and room as in the previous section, and the conditions of Proposition 4.2 are satisfied, then $\boldsymbol{H}$ can be expected to have full row rank with probability 1.

Proposition 4.2. *The following conditions are necessary for perfect reconstruction of message signals at the listeners.*

- *(a) The number of rows of* $\boldsymbol{H}$ *should be at least as large as the length of* $\boldsymbol{y}_{in} \implies (L_x + L_h - 1) \geq N$.
- *(b) There should be at least as many columns as rows in* $\boldsymbol{H}$.
- *(c)* L_x *needs to be greater than the highest relative time delay among each listener-speaker pair.*

Proof. (a) ensures that we have sufficient samples to generate the desired message length; (b) is elementary linear algebra; (c) ensures that "silent" regions do not exist within a signal generated at a listening point. □

It should be noted that both of our approaches satisfy the condition in (a) with equality. Also, (b) gives a lower bound on the number of speakers, L, needed for reconstruction, i.e., $L \geq \frac{NK}{L_x}$. This is lower than the number of speakers needed by the MCCS approach, as per Proposition 4.1

5. EXPERIMENTAL RESULTS

We evaluate the performance of the two proposed techniques using both numerical and real experiments. The numerical experiments are performed with 6 loudspeakers randomly placed in a simulated convex room of size 7 m × 8 m having walls with absorption coefficient 0.35. RIRs between the speakers and listeners are calculated based on image source model, using the `pyroomacoustics` package [24]. We perform the real experiments in an office space of size 10 m × 6 m using two Genelec 8030B and four Genelec 8010A loudspeakers. The RIRs are measured using the exponential sine sweep technique [25]. In all experiments, the power of signals emitted by the loudspeakers is kept fixed. The intelligibility of the generated sounds is assessed using Short-Time Objective Intelligibility (STOI) [26] measure.

5.1. Numerical experiments

5.1.1. Perfect reconstruction: A case for echoes

In order to provide insight into the importance of echoes in our solution, we first perform an experiment in a simulated anechoic room. We randomly place two listeners inside the room and calculate STOI scores of the signals arriving there using the two approaches. An additional location is randomly chosen to examine the signal degradation outside the target focusing spots. We then repeat the same experiment but in the presence of echoes. Fig. 1 (a) shows that in the anechoic setting, while the signal at the first listener has high intelligibility with STOI scores close to 1 for both approaches, the second listener does not. On the other hand, Fig. 1(b) shows that in the presence of echoes, signal intelligibility is restored at the second listener as well. This indicates that the spatial diversity provided by echoes helps in conditioning the channel matrix $\boldsymbol{H}$, which in turn supports perfect reconstruction of messages at target locations.

5.1.2. Signal degradation outside focusing spots

Both Fig. 1 (a) and (b) indicate that the nullspace-based method has a greater impact on signal degradation at the location chosen outside the focusing spots. To examine this further, we calculate STOI scores at 4200 locations in a simulated reverberant room and create heat maps as shown in Fig. 1 (c) and (d). In both plots, the bright spots at the locations of intended listeners indicate high intelligibility. However, regions outside the focusing spots in Fig. 1 (d) have relatively lower STOI scores as compared to Fig. 1 (c), thus indicating towards better jamming capabilities of the nullspace approach.

Both methods perform signal degradation outside the focusing spots using noise. To understand how these random signals result

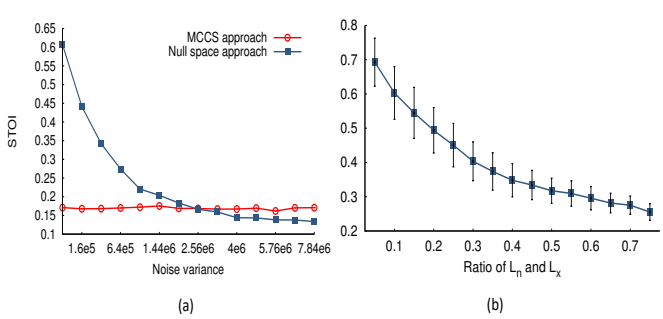

Fig. 2: (a) STOI vs noise variance for the 2 methods outside focusing spots. (b) STOI vs noise length as a proportion of overall input length for MCCS approach outside focusing spots.

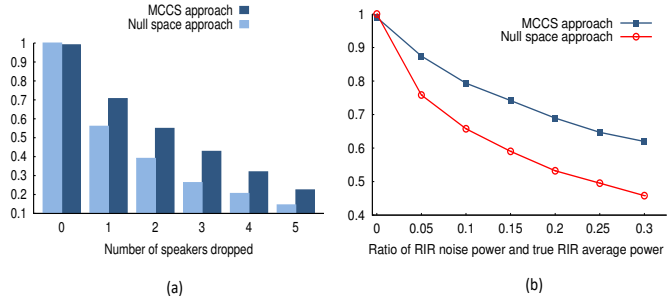

Fig. 3: Robustness analysis. Impact of (a) speaker malfunction and (b) inaccuracies in RIR estimates on STOI scores at focusing spots.

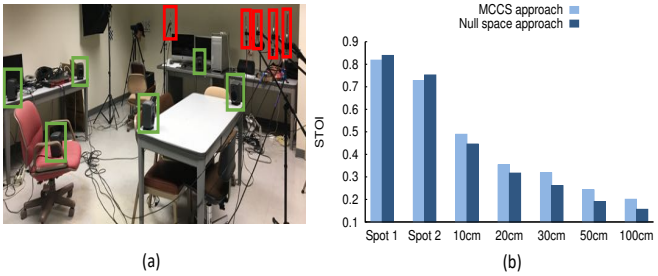

Fig. 4: (a) Experimental setup: speakers represented in green, and microphones in red boxes. (b) STOI values measured at two focusing spots, and at different distances from Spot 2 in a real room setting.

in unintelligibility of sound, we first investigate the role of noise variance. For 100 randomly selected speaker-listener configurations, we check the impact of increasing noise variance on STOI values for both methods. Fig. 2 (a) shows a decline in median STOI scores as the input noise power is increased for the nullspace approach, whereas they do not change much for the MCCS method.

This result is not surprising because in the nullspace approach, noise is fed into the loudspeakers with the message signals in an additive sense. Thus, a deterioration of SNR and subsequent STOI decline is expected with increase in noise variance. However, the signal emitted by the i^{th} loudspeaker is $\boldsymbol{x}_i = \boldsymbol{n}_i * \boldsymbol{g}_i$ for MCCS method. Here, if the variance of $\boldsymbol{n}_i$ is increased, $\boldsymbol{g}_i$ simply gets scaled to preserve the original $\boldsymbol{x}_i$.

We now investigate the factors that impact the jamming capability of the MCCS approach. Recall that this method involves "scrambling" of message-carrying input filters $\boldsymbol{g}_i$ by noise which are thereby appropriately descrambled at the intended locations by the correct RIR values. Thus, we expect that longer noise vectors would have a stronger impact on signal integrity when the RIR changes. To verify this claim, we vary the length of noise vectors L_{n} as a proportion of a fixed length L_{x}, and calculate the STOI scores for 100 randomly chosen speaker-listener configurations. Fig. 2(b) verifies that increasing the length of noise vectors leads to a decrease in median intelligibility scores outside the focusing spots.

These results point towards an interesting phenomenon. Given unlimited available input power at the speakers, one could arbitrarily improve jamming by increasing noise power in the nullspace method. However, in MCCS approach, an arbitrary increase in jamming by increasing L_{n} is not feasible, because for a fixed message length N and fixed L_{h}, $L_{\text{x}} = L_{\text{g}} + L_{\text{n}} - 1$ is fixed, and one can only increase L_{n}, as long as $L_{\text{g}} \geq \frac{NK}{L}$ (from Proposition 4.1).

5.1.3. Robustness to system failures and uncertainties

We assess how the reconstruction of audio messages at the target listeners is affected by system failures and uncertainties: (i) malfunction of loudspeakers while emitting audio signals, and (ii) errors in RIR measurements. We did simulations over 100 random speaker–listener configurations and examined the behavior of the STOI scores. In (i), we compute the appropriate $\mathbf{x}_i$ (to be emitted by the i^{th} loudspeaker) for a system of 6 speakers. However, while measuring STOI at the listeners, not all speakers are used. Fig. 3(a) shows that the STOI scores decline as more speakers are dropped, and the decline is more rapid for the nullspace method as compared to MCCS approach.

On the other hand, we analyze the robustness to channel measurement errors by computing $\mathbf{x}_i$ using RIR values with white Gaussian noise added to them. These erroneous $\mathbf{x}_i$ are then convolved with the true RIRs to compute the signals arriving at focusing spots. Fig. 3(b) indicates that errors in the knowledge of RIRs before signal transmission by the loudspeakers lead to reduced intelligibility at the focusing spots. Again, the MCCS approach shows more robustness to uncertainties as compared to the nullspace approach.

5.2. Experiment in a real setting

We perform an experiment to evaluate the two approaches in a real room with 6 loudspeakers and measure the STOI scores of generated sounds with microphones at 7 locations. The experimental setup is shown in Fig. 4 (a). Two microphones are chosen to be the focusing spots, and the rest are placed at increasing distances from Spot 2. Fig. 4 (b) shows the measured STOI values. The observed intelligibility at the two spots is good with high STOI scores, and the signals become considerably degraded 50 cm away from the focusing spots. As expected from simulations, the nullspace approach has a stronger impact on signal degradation outside the target listeners.

6. CONCLUSION

We present two approaches to address the private audio communication problem in a reverberant room. Both approaches are based on emitting noise signals from loudspeakers and then utilizing echoes in the room to ensure that they yield intelligible messages at selected locations, while being incoherent elsewhere. Simulated and real experiments suggest that with just 6 loudspeakers and a few impulse response measurements, we can deliver clear audio messages at the desired locations while ensuring unintelligibility everywhere else.

7. REFERENCES

[1] M. Poletti, "An investigation of 2-d multizone surround sound systems," in *125th Audio Engineering Society Convention*, Oct 2008.

[2] Y. J. Wu and T. D. Abhayapala, "Spatial multizone soundfield reproduction: Theory and design," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 6, pp. 1711–1720, Aug 2011.

[3] T. Betlehem, W. Zhang, M. A. Poletti, and T. D. Abhayapala, "Personal sound zones: Delivering interface-free audio to multiple listeners," *IEEE Signal Processing Magazine*, vol. 32, no. 2, pp. 81–91, March 2015.

[4] S. J. Elliott, J. Cheer, J. Choi, and Y. Kim, "Robustness and regularization of personal audio systems," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 7, pp. 2123–2133, Sep. 2012.

[5] Y. Cai, M. Wu, and J. Yang, "Sound reproduction in personal audio systems using the least-squares approach with acoustic contrast control constraint," *The Journal of the Acoustical Society of America*, vol. 135, no. 2, pp. 734–741, 2014. [Online]. Available: https://doi.org/10.1121/1.4861341

[6] J.-W. Choi and Y.-H. Kim, "Generation of an acoustically bright zone with an illuminated region using multiple sources," *The Journal of the Acoustical Society of America*, vol. 111, no. 4, pp. 1695–1700, 2002.

[7] A. J. Berkhout, D. de Vries, and P. Vogel, "Acoustic control by wave field synthesis," *The Journal of the Acoustical Society of America*, vol. 93, no. 5, pp. 2764–2778, 1993.

[8] D. B. Ward and T. D. Abhayapala, "Reproduction of a plane-wave sound field using an array of loudspeakers," *IEEE Transactions on speech and audio processing*, vol. 9, no. 6, pp. 697–707, 2001.

[9] W. Jin, W. B. Kleijn, and D. Virette, "Multizone soundfield reproduction using orthogonal basis expansion," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2013, pp. 311–315.

[10] Y. Liu, J. Casebeer, and I. Dokmanić, "Cocktails, but no party: Multipath-enabled private audio," in *16th International Workshop on Acoustic Signal Enhancement (IWAENC)*, Sep. 2018, pp. 186–190.

[11] R. Negi and S. Goel, "Secret communication using artificial noise," in *62nd IEEE Vehicular Technology Conference*, vol. 3, Sep. 2005, pp. 1906–1910.

[12] S. Goel and R. Negi, "Guaranteeing secrecy using artificial noise," *IEEE Transactions on Wireless Communications*, vol. 7, no. 6, pp. 2180–2189, June 2008.

[13] J. Donley, C. Ritz, and W. B. Kleijn, "Improving speech privacy in personal sound zones," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2016, pp. 311–315.

[14] C. E. Shannon, "Communication theory of secrecy systems," *The Bell System Technical Journal*, vol. 28, no. 4, pp. 656–715, Oct 1949.

[15] I. Csiszar and J. Korner, "Broadcast channels with confidential messages," *IEEE Transactions on Information Theory*, vol. 24, no. 3, pp. 339–348, May 1978.

[16] A. D. Wyner, "The wire-tap channel," *Bell System Technical Journal*, vol. 54, no. 8, pp. 1355–1387, 1975. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/j.1538-7305.1975.tb02040.x

[17] S. Goel and R. Negi, "Secret communication in presence of colluding eavesdroppers," in *IEEE Military Communications Conference*, vol. 3, Oct 2005, pp. 1501–1506.

[18] J. Barros and M. R. D. Rodrigues, "Secrecy capacity of wireless channels," in *IEEE International Symposium on Information Theory*, July 2006, pp. 356–360.

[19] A. Mukherjee and A. L. Swindlehurst, "Detecting passive eavesdroppers in the mimo wiretap channel," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2012, pp. 2809–2812.

[20] A. Chaman, J. Wang, J. Sun, H. Hassanieh, and R. Roy Choudhury, "Ghostbuster: Detecting the presence of hidden eavesdroppers," in *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*. ACM, 2018, pp. 337–351.

[21] C. Stagner, A. Conrad, C. Osterwise, D. G. Beetner, and S. Grant, "A practical superheterodyne-receiver detector using stimulated emissions," *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no. 4, pp. 1461–1468, April 2011.

[22] W. Jin and W. B. Kleijn, "Theory and design of multizone soundfield reproduction using sparse methods," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 12, pp. 2343–2355, Dec 2015.

[23] T. Betlehem and T. D. Abhayapala, "Theory and design of sound field reproduction in reverberant rooms," *The Journal of the Acoustical Society of America*, vol. 117, no. 4, pp. 2100–2111, 2005. [Online]. Available: https://doi.org/10.1121/1.1863032

[24] R. Scheibler, E. Bezzam, and I. Dokmanić, "Pyroomacoustics: A python package for audio room simulation and array processing algorithms," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2018, pp. 351–355.

[25] A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *108th Audio Engineering Society Convention*, Feb 2000. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=10211

[26] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "A short-time objective intelligibility measure for time-frequency weighted noisy speech," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, March 2010, pp. 4214–4217.