IMMERSIVE ENHANCEMENT AND REMOVAL OF LOUDSPEAKER SOUND USING WIRELESS ASSISTIVE LISTENING SYSTEMS AND BINAURAL HEARING DEVICES

Ryan M. Corey

Discovery Partners Institute
University of Illinois Chicago

Andrew C. Singer

University of Illinois Urbana-Champaign

ABSTRACT

Wireless assistive listening devices (ALDs), such as induction loops, radio-frequency transmitters, and digital streaming systems, improve accessibility for people with hearing loss by transmitting from a venue's sound system directly to the listener. Today, ALDs are used primarily for lectures and performances. When paired with advanced hearing devices, however, they could form part of an augmented listening system that lets users "remix" sounds in their environment, including from loudspeakers in public spaces. For example, users could amplify public announcements or suppress background music while having a conversation. In the proposed system, a binaural adaptive filter uses the ALD signal to estimate the loudspeaker sound at the ears. The hearing device can then either enhance or remove the loudspeaker sound in the hearing device output while preserving other nearby sounds. We demonstrate the proposed system using several commercial ALDs and assess the effects of delay, bandwidth, distortion, and noise on real-world system performance.

Index Terms— Hearing aids, assistive listening device, adaptive filtering, binaural signal processing

1. INTRODUCTION

Augmented listening systems, such as hearing aids, advanced earphones, and augmented reality devices, alter and enhance the sounds that humans hear [1, 2]. Augmented listening can be viewed as "remixing" real-world sounds [3], for example to amplify sounds of interest and suppress unwanted noise. Such systems require robust real-time source separation [4, 5]. They must also be perceptually transparent, processing sound with no more than a few milliseconds of delay [6,7] and preserving spatial cues such as interaural time and level differences [8]. For example, researchers have proposed binaural beamformers that enhance a target sound while preserving the interaural cues of both target and background sounds [9–13].

Because single-device source separation can be difficult in complex environments, augmented listening systems can benefit from external devices such as microphones worn by talkers. Remote microphones have been shown to reliably improve intelligibility in adverse conditions [14–16]. In an immersive binaural system, remote microphone signals can be used to calibrate a binaural beamformer [17,18] or spatialized using either direction-of-arrival modeling [19] or binaural adaptive filtering [20]. In this work, we consider another source that users might want to remix: sound played over a loudspeaker system. Users could amplify public addresses in a train station or turn down distracting background music in a restaurant while still listening to conversation partners.

This research was supported by the National Science Foundation under Grant No. 1919257 and by an appointment to the Intelligence Community Postdoctoral Research Fellowship Program at the University of Illinois Urbana-Champaign, administered by Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and the Office of the Director of National Intelligence.

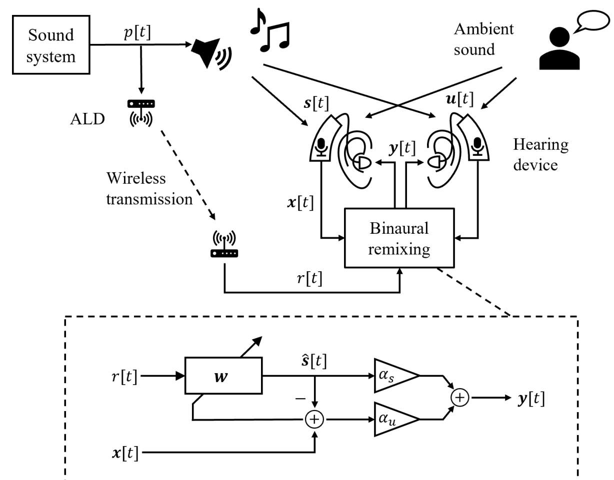

Fig. 1. An immersive remixing system allows users personalize the level of sound played over a loudspeaker. A wireless ALD broadcasts the loudspeaker signal to a listening device, which either enhances or suppresses it while preserving other ambient sounds.

With loudspeaker sound, unlike live speech and other natural sources, a remixing system can access the noise-free source signal; it need only be transmitted from the sound system to the hearing device. Many classrooms, theaters, and places of worship already use audio broadcast systems, known as assistive listening devices (ALDs), to improve accessibility for people with hearing loss. An ALD transmits the signal being played over the venue's loudspeakers to one or more users [21, 22]. Analog frequency-modulation (FM) and infrared broadcast systems offer low latency and wide bandwidth, but require users to borrow dedicated receivers. Some hearing-loss advocates prefer hearing loops, which use electromagnetic induction to produce sound directly in compatible hearing devices [23]. Newer systems use WiFi to stream sound to the user's smartphone, while an upcoming Bluetooth one-to-many broadcast protocol is expected to allow streaming directly to hearing devices as well as consumer earphones [24]. These low-cost digital broadcast technologies could allow ALDs to be installed not just in large performance venues but in nearly any space with a sound system.

In anticipation of widespread, high-quality ALDs, we consider the design of an immersive remixing system that integrates a wireless ALD signal with the sound captured by a binaural hearing device. Most current ALDs transmit monaural signals. To perform real-time cancellation or to enhance the broadcast sound while preserving spatial cues, the wireless signal must be aligned in time and spectrum with the live sound at the ears. This work applies the adaptive filtering approach of [20] to the ALD signal. A pair of adaptive filters [25] processes the transmitted signal to match the sound at the ears and updates the acoustic channel estimate as the listener moves. The ALD has access to a noise-free copy of the source signal, but the wireless transmission can introduce delay, distortion, noise, and packet loss that harm the performance of the system. We demonstrate the proposed remixing system using real-room recordings with a moving human listener and several commercial ALDs. The results show that transmission methods that sound good to human listeners are not necessarily suitable for adaptive filtering.

2. ADAPTIVE FILTER FOR SOUND REMIXING

2.1. System model

Consider the system shown in Figure 1 over a short time scale during which the listener can be considered stationary. A loudspeaker plays the discrete-time signal $p[t]$, which propagates to the two earpiece microphones according to the unit pulse response vector $\mathbf{a}_s[\tau] \in \mathbb{R}^2$. The earpieces also capture ambient sound $\mathbf{u}[t] \in \mathbb{R}^2$, which includes all sources except the loudspeaker. The sound at the ears is

$$\mathbf{x}[t] = (\mathbf{a}_s \star p)[t] + \mathbf{u}[t] \tag{1}$$
$$= \mathbf{s}[t] + \mathbf{u}[t], \tag{2}$$

where $\mathbf{s}[t] = (\mathbf{a}_s \star p)[t]$ is the loudspeaker sound at the two ears and $\star$ denotes linear convolution.

The loudspeaker signal is also transmitted from an ALD to the hearing device. The received reference signal is

$$r[t] = (a_r \star p)[t] + z[t], \tag{3}$$

where $a_r[t]$ is the unit pulse response of the wireless channel and $z[t]$ is additive noise.

The goal of immersive remixing is to alter the relative levels of the loudspeaker sound and other ambient sounds while preserving timing and spatial cues. Let α_s and α_u be nonnegative target gains for the loudspeaker and ambient sounds, respectively. The desired output of the hearing device is given by

$$\mathbf{d}[t] = \alpha_u \mathbf{u}[t] + \alpha_s \mathbf{s}[t] \tag{4}$$
$$= \alpha_u \mathbf{x}[t] + (\alpha_s - \alpha_u)\mathbf{s}[t]. \tag{5}$$

The binaural earpiece output $\mathbf{y}[t] \in \mathbb{R}^2$ is obtained by mixing the live sound and a filtered version of the reference sound:

$$\mathbf{y}[t] = \alpha_u \mathbf{x}[t] + (\alpha_s - \alpha_u)\hat{\mathbf{s}}[t], \tag{6}$$

where $\hat{\mathbf{s}}[t]$ is the estimated loudspeaker sound given by

$$\hat{\mathbf{s}}[t] = (\mathbf{w} \star r)[t], \tag{7}$$

and $\mathbf{w}[\tau] \in \mathbb{R}^2$ is a finite-impulse-response filter. The error is

$$\boldsymbol{\epsilon}[t] = \mathbf{d}[t] - \mathbf{y}[t] \tag{8}$$
$$= (\alpha_s - \alpha_u)(\mathbf{s}[t] - \hat{\mathbf{s}}[t]). \tag{9}$$

Notice that as long as the estimator behaves linearly, the ambient sound $\mathbf{u}[t]$ is processed without error and the overall output error is from estimating the loudspeaker sound $\mathbf{s}[t]$.

The filter coefficients should be chosen to minimize the mean square error $\mathbb{E}\left\|\mathbf{s} - \hat{\mathbf{s}}\right\|^2$. However, since we cannot observe $\mathbf{s}$ directly, the filter instead adaptively minimizes the error between its output and the earpiece mixture $\mathbf{x}$:

$$\mathbf{w} = \arg\min \mathbb{E}\left\|\mathbf{x} - \hat{\mathbf{s}}\right\|^2. \tag{10}$$

In our experiments, $\mathbf{w}$ is adapted over time using the normalized least-mean-squares algorithm with first-order pre-whitening [26]:

$$\vec{\mathbf{w}}^{t+1} \leftarrow \vec{\mathbf{w}}^{t} + \bar{\mu}\frac{(\mathbf{x}[t] - \hat{\mathbf{s}}[t])\vec{r}[t]}{\left\|\vec{r}[t]\right\|^2}, \tag{11}$$

where $\vec{\mathbf{w}} = [\mathbf{w}[0], \mathbf{w}[1], \dots]$, $\vec{r} = [r[t], r[t-1], \dots]$, and $0 < \bar{\mu} \ll 1$.

Note that for simplicity, our signal model only considers the earpiece inputs and outputs and does not account for the direct acoustic path to the ear. If at least one sound is to be suppressed rather than enhanced, then the hearing device should provide passive sound isolation or active noise cancellation.

2.2. Stationary solution

If the listener remains still long enough for the adaptive filter to converge, then it will reach the minimum-mean-square-error (MMSE) solution that solves (10). Suppose that $p[t]$ and $z[t]$ are wide-sense stationary random processes that are uncorrelated with each other and with $\mathbf{u}[t]$ and that have power spectral densities $R_p(\omega)$ and $R_z(\omega)$, respectively. Let $A_r(\omega)$ and $\mathbf{A}_s(\omega)$ be the discrete-time Fourier transforms of $a_r[t]$ and $\mathbf{a}_s[t]$, respectively. Let $R_r(\omega)$ be the power spectral density of $r[t]$ and let $\mathbf{R}_{\mathbf{x},r}(\omega)$ be the cross-spectrum between $\mathbf{x}[t]$ and $r[t]$. In the frequency domain, the (possibly noncausal) MMSE filter is given by

$$\mathbf{W}_{\mathrm{MMSE}}(\omega) = \mathbf{R}_{\mathbf{x},r}(\omega)R_r^{-1}(\omega) \tag{12}$$
$$= \frac{\mathbf{A}_s(\omega)R_p(\omega)A_r^*(\omega)}{A_r(\omega)R_p(\omega)A_r^*(\omega) + R_z(\omega)} \tag{13}$$
$$= \frac{|A_r(\omega)|^2 R_p(\omega)}{|A_r(\omega)|^2 R_p(\omega) + R_z(\omega)} \cdot \frac{\mathbf{A}_s(\omega)}{A_r(\omega)}. \tag{14}$$

The first term is a scalar Wiener filter that suppresses additive noise in the broadcast channel. The second is the relative transfer function between the earpieces and reference signal. Because the filter models the relationship between $r[t]$ and $\mathbf{s}[t]$, $\mathbf{W}_{\mathrm{MMSE}}$ does not depend on $\mathbf{u}[t]$.

The filter expression illustrates the impact that the ALD has on the performance of the remixing system. With a high-quality ALD, $A_r(\omega) \approx 1$ and $R_z(\omega) \ll R_p(\omega)$ so that $\mathbf{W}_{\mathrm{MMSE}}(\omega) \approx \mathbf{A}_s(\omega)$. That is, the filter approximates the transfer function from the loudspeaker to the ears. If the broadcast channel introduces delay that is longer than the direct acoustic path, then $\mathbf{A}_s/A_r$ is noncausal and the hearing device must delay its output. If the broadcast channel is not spectrally flat, or if its signal-to-noise ratio varies across frequency, then the adaptive filter must perform frequency-dependent equalization, which may incur additional delay.

2.3. Remixing gains and system performance

The performance of the system depends on the accuracy of the estimated filter and on the choice of remixing gains α_s and α_u. If $\alpha_s = \alpha_u$, then the system trivially achieves perfect performance by reproducing the signal captured by the earpiece microphones; the adaptive filter is not used at all.

If $\alpha_s = 0$, then the system attempts to fully cancel the loudspeaker signal from the earpiece signal. Signal cancellation is highly sensitive to the accuracy of the filter. If $\mathbf{w}$ is the MMSE filter (14), then the deviation between the modeled and the true acoustic transfer function is

$$\mathbf{W}(\omega)A_r(\omega) - \mathbf{A}_s(\omega) = \frac{R_z(\omega)}{|A_r(\omega)|^2 R_z(\omega) + R_z(\omega)} \mathbf{A}_s(\omega). \quad (15)$$

Thus, we can expect weaker cancellation performance at frequencies where the broadcast signal is noisy.

If $\alpha_s > \alpha_u$, then the goal is to enhance the loudspeaker sound. In most applications, it is not necessary to perfectly match the acoustic impulse response, but the filter does need to preserve spatial cues including interaural level differences (ILD) and interaural time/phase differences (ITD/IPD). Both can be derived from the interaural transfer function (ITF). If $\mathbf{s}[t]$ has discrete-time Fourier transform $\mathbf{S}(\omega)$, then the input ITF is given by

$$\mathrm{ITF}_s^{\mathrm{in}}(\omega) = \frac{\mathbf{e}_2^T \mathbf{S}(\omega)}{\mathbf{e}_1^T \mathbf{S}(\omega)} = \frac{\mathbf{e}_2^T \mathbf{A}_s(\omega)}{\mathbf{e}_1^T \mathbf{A}_s(\omega)}, \quad (16)$$

where $\mathbf{e}_1^T = [1, 0]$ and $\mathbf{e}_2^T = [0, 1]$ are unit vectors that select the left- and right-ear signals. The ILD and IPD are the magnitude and phase, respectively, of the ITF.

From (6), it can be shown that the output ITF is given by

$$\mathrm{ITF}_s^{\mathrm{out}}(\omega) = \frac{\mathbf{e}_2^T \left(\mathbf{A}_s(\omega) + \frac{\alpha_u - \alpha_s}{\alpha_s} \left(\mathbf{A}_s(\omega) - \mathbf{W}(\omega)A_r(\omega)\right)\right)}{\mathbf{e}_1^T \left(\mathbf{A}_s(\omega) + \frac{\alpha_u - \alpha_s}{\alpha_s} \left(\mathbf{A}_s(\omega) - \mathbf{W}(\omega)A_r(\omega)\right)\right)}. \quad (17)$$

For the MMSE filter, $\mathbf{W}(\omega)$ is parallel to $\mathbf{A}_s(\omega)$, so $\mathrm{ITF}_s^{\mathrm{out}}(\omega) = \mathrm{ITF}_s^{\mathrm{in}}(\omega)$ regardless of the noise level. If the filter has not converged correctly to the MMSE solution, then the spatial cue distortion depends on both the discrepancy between $\mathbf{W}(\omega)$ and $\mathbf{A}_s(\omega)$ and on the relative gain ratio $(\alpha_u - \alpha_s)/\alpha_s$. The more the system tries to amplify the loudspeaker sound relative to the ambient sound, the more sensitive it is to spatial errors.

3. PERFORMANCE WITH COMMERCIAL WIRELESS SYSTEMS

Although the signal being played over the loudspeaker is known exactly, it must be transmitted over a wireless system that may not perform reliably. In this section, we assess the feasibility of current commercial wireless ALDs for the immersive remixing system and consider the requirements that future systems should meet.

All experiments were performed in an acoustically treated laboratory ($T_{60} \approx 150$ ms). One loudspeaker played orchestral music that was simultaneously broadcast over an ALD. The ambient sound was speech from the VCTK corpus [27] played over a second loudspeaker. A binaural behind-the-ear hearing device was simulated by a pair of omnidirectional lavalier microphones on a human subject. To quantify separation performance, the music and speech were recorded separately and mixed in software. The adaptive system was tested on a one-minute excerpt of the recorded data downsampled to 24 kHz. The filters had length 512 samples, or about 21 ms. Unless otherwise specified, the adaptive filters used $\bar{\mu} = 0.01$.

3.1. Commercial assistive listening devices

The experiments compared four commercial assistive listening devices, summarized in Table 1. The receiver output signal from each

System	Delay (ms)	Distortion
Wire	0	–
Advanced FM	1	–
Basic FM	1	Spectral slope
Induction loop	1	Additive noise
WiFi	500–700	Packet loss, delay drift

Table 1. Commercial assistive listening devices

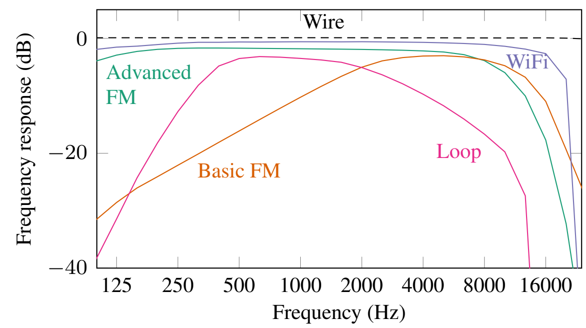

Fig. 2. Measured frequency responses of several commercial wireless ALDs. Absolute levels are arbitrary and the curves are vertically positioned for legibility.

system was recorded synchronously by the studio interface during playback. The frequency responses of the systems, shown in Fig. 2, were measured using frequency sweeps. As an ideal reference, one input was hard-wired to the playback output.

We compared two FM broadcast systems: one high-end system (Williams AV "FM+ PRO") and one entry-level system (Williams AV "Personal FM Value Pack"). Both systems had negligible delay and additive noise. The advanced FM system has a flat frequency response across its rated bandwidth of 16 kHz. Both transmitters apply a 6 dB/octave pre-emphasis filter; however, the basic FM receiver evidently does not apply a corresonding de-emphasis filter, causing spectral distortion in its output.

The advanced FM system also includes WiFi streaming (Williams AV WaveCAST) that broadcasts a 30 kbps compressed audio stream to a smartphone app. The performance of this system was found to depend strongly on network quality, with poor connections causing frequent dropouts. After careful network optimization, there were few dropped frames in our experimental recordings. However, the received signal exhibited a large and time-varying delay. The measured delay increased linearly over time, consistent with time expansion by 500 parts per million.

Induction loops are normally installed within the floor of an auditorium or other venue. For this laboratory experiment, we used a tabletop loop system (Williams AV PLA 90) and a telecoil receiver (OtoJoy LoopBuds) positioned less than one meter in front of the transmitter. Unlike the other tested ALDs, the loop-telecoil system introduced noticeable noise, including 60 Hz electrical hum.

3.2. Enhancement performance with different ALDs

To understand the effects of different channel distortions, we compared the performance of the ALD reference signals for an enhancement task with a nonmoving listener. To amplify the loudspeaker sound without perceptible distortion or echoes, the adaptive filter

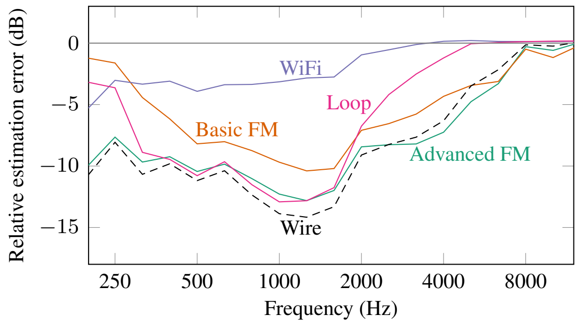

Fig. 3. Square-error performance (lower is better) of the adaptive filter for a nonmoving listener using different ALDs. Curves show the error power $|\hat{s} - s|^2$ relative to $|s|^2$ averaged over the two ears.

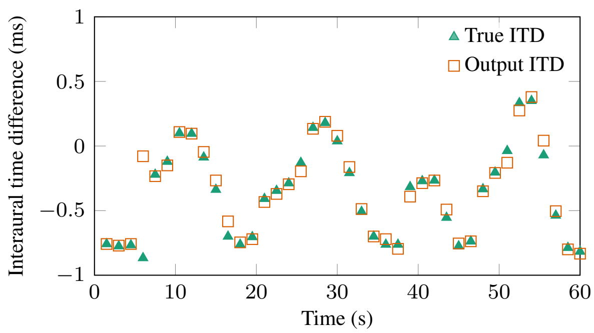

Fig. 4. Interaural time differences for music played over a loudspeaker at the input and output of the proposed enhancement system.

much match its output to the signal at the two ears. Figure 3 shows the mean square error of the estimated loudspeaker sound $\hat{\mathbf{s}}[t]$ relative to the true earpiece recording $\mathbf{s}[t]$.

The advanced FM system performs nearly as well as the direct wired connection. The loop performed worse at higher frequencies, where it had limited bandwidth compared to the FM systems. The basic FM system had reduced performance at low frequencies, where the pre-emphasis filter removes signal information. When the experiment was repeated with a digital de-emphasis filter, low-frequency performance improved but was still worse than the advanced system.

The WiFi system—ostensibly the most modern technology—presented the greatest challenges. The 500 ms delay exceeds the acoustic time of flight in many spaces, making it unsuitable for live sound. To allow a fair comparison, the recorded WiFi signal was advanced by 500 ms relative to the others. However, the adaptive filter was unable to adapt to the 500 parts-per-million delay drift, even when the filter was lengthened to accomodate the full range of delays during the sample. When the signal was resampled using the interpolation method of [28], the performance was comparable to that of the analog systems. Although the small drift might not be noticeable to human listeners, it is detrimental to the adaptive filter.

3.3. Preservation of spatial cues for a moving listener

One of the advantages of the proposed immersive remixing system compared to a conventional ALD is that it preserves the listener's spatial awareness. In this experiment, the human subject walked around the room during the experiment, including moving toward and away from the loudspeaker and turning their head relative to the sound sources. Figure 4 shows the interaural time difference of the unprocessed music at the ears and of the spatialized signal at the filter output, which used the advanced FM signal as a reference. The output ITDs closely match the true ITDs, showing that the adaptive filter is able to track the user's motion.

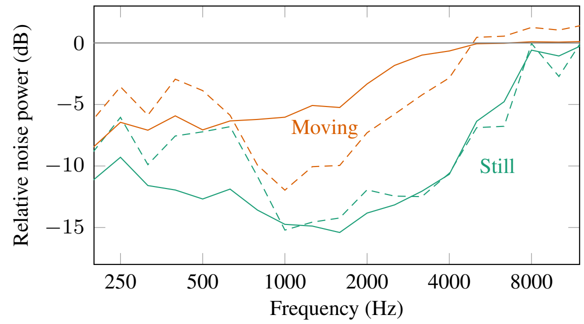

Fig. 5. Residual music power at the output (lower is better), relative to the input, of the loudspeaker cancellation system. The solid curves used $\bar{\mu} = 0.01$ and the dashed curves used $\bar{\mu} = 0.1$.

3.4. Cancellation performance with a moving listener

A novel application of the proposed system is to remove nuisance sounds, such as background music, while preserving other sounds, such as speech. Figure 5 shows the relative suppression of the unwanted music signal for different motion conditions and step sizes. A smaller step size is better when the listener is not moving. When the listener moves, the filter must adapt more quickly to keep up, and a larger step size allows it to cancel more of the music.

However, the faster-adapting filter also behaves nonlinearly and cancels some of the speech signal. With $\bar{\mu} = 0.1$, intelligible speech can be heard in the filter output $\hat{\mathbf{s}}[t]$, even though its input is the music $r[t]$. This modulation behavior violates the linearity and statistical assumptions used to analyze the filter in Section 2 and limits the practicality of the proposed method in dynamic scenarios. Further research is required to find techniques that can quickly track changes in the channel while avoiding nonlinear distortion.

4. CONCLUSIONS

The experiments show that it is feasible to combine an ALD with a hearing device to immersively remix loudspeaker sound, even with a moving listener. The system can even cancel unwanted noise playing over a sound system, effectively repurposing ALDs for the opposite of their intended function. However, the system relies on a high-quality reference signal. Distortions that might not bother a human listener, like a pre-emphasis filter or slight time warping, are detrimental to the adaptive filter. The most important requirement for an ALD to be used for immersive remixing is a low and stable delay. Further research is required to develop adaptive systems that are robust against lost packets and other timing discrepancies in live digital audio streams. Future ALDs should also support multichannel broadcasts so that an augmented listening system can track sound from multiple loudspeakers and preserve the intended stereo image. If next-generation wireless systems are designed with augmented listening in mind, then they can work alongside remote microphones, microphone arrays, and other connected acoustic sensors to give users unprecedented control over their listening experience.

5. REFERENCES

[1] Vesa Valimaki, Andreas Franck, Jussi Ramo, Hannes Gamper, and Lauri Savioja, "Assisted listening using a headset: Enhancing audio perception in real, augmented, and virtual environments," *IEEE Signal Processing Magazine*, vol. 32, no. 2, pp. 92–99, 2015.

[2] Ryan M. Corey, *Microphone Array Processing for Augmented Listening*, Ph.D. thesis, University of Illinois Urbana-Champaign, 2019.

[3] Ryan M. Corey and Andrew C. Singer, "Binaural audio source remixing with microphone array listening devices," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020.

[4] Emmanuel Vincent, Tuomas Virtanen, and Sharon Gannot, *Audio Source Separation and Speech Enhancement*, Wiley, 2018.

[5] Shoji Makino, Ed., *Audio Source Separation*, Springer, 2018.

[6] Jeremy Agnew and Jeffrey M. Thornton, "Just noticeable and objectionable group delays in digital hearing aids," *Journal of the American Academy of Audiology*, vol. 11, no. 6, pp. 330–336, 2000.

[7] Michael A. Stone, Brian C.J. Moore, Katrin Meisenbacher, and Ralph P. Derleth, "Tolerable hearing aid delays. V. Estimation of limits for open canal fittings," *Ear and hearing*, vol. 29, no. 4, pp. 601–617, 2008.

[8] Jens Blauert, *Spatial hearing: The psychophysics of human sound localization*, MIT press, 1997.

[9] Simon Doclo, Thomas J. Klasen, Tim Van den Bogaert, Jan Wouters, and Marc Moonen, "Theoretical analysis of binaural cue preservation using multi-channel Wiener filtering and interaural transfer functions," in *International Workshop on Acoustic Echo and Noise Control (IWAENC)*, 2006.

[10] Tim Van den Bogaert, Simon Doclo, Jan Wouters, and Marc Moonen, "Speech enhancement with multichannel Wiener filter techniques in multimicrophone binaural hearing aids," *The Journal of the Acoustical Society of America*, vol. 125, no. 1, pp. 360–371, 2009.

[11] Daniel Marquardt, Volker Hohmann, and Simon Doclo, "Interaural coherence preservation in multi-channel Wiener filtering-based noise reduction for binaural hearing aids," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 23, no. 12, pp. 2162–2176, 2015.

[12] Elior Hadad, Simon Doclo, and Sharon Gannot, "The binaural LCMV beamformer and its performance analysis," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 24, no. 3, pp. 543–558, 2016.

[13] Elior Hadad, Simon Doclo, Sven Nordholm, and Sharon Gannot, "A class of Pareto optimal binaural beamformers," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 30, pp. 2612–2628, 2022.

[14] Arthur Boothroyd, "Hearing aid accessories for adults: The remote FM microphone," *Ear and Hearing*, vol. 25, no. 1, pp. 22–33, 2004.

[15] Linda Thibodeau, "Comparison of speech recognition with adaptive digital and FM remote microphone hearing assistance technology by listeners who use hearing aids," *American Journal of Audiology*, vol. 23, no. 2, pp. 201–210, 2014.

[16] Jace Wolfe, Mila Morais Duke, Erin Schafer, Christine Jones, Hans E. Mülder, Andrew John, and Mary Hudson, "Evaluation of performance with an adaptive digital remote microphone system and a digital remote microphone audio-streaming accessory system," *American Journal of Audiology*, vol. 24, no. 3, pp. 440–450, 2015.

[17] Nico Gößling and Simon Doclo, "Relative transfer function estimation exploiting spatially separated microphones in a diffuse noise field," in *International Workshop on Acoustic Signal Enhancement (IWAENC)*, 2018.

[18] Nico Gößling, *Binaural Beamforming Algorithms and Parameter Estimation Methods Exploiting External Microphones*, Ph.D. thesis, University of Oldenburg, 2020.

[19] James M. Kates, Kathryn H. Arehart, Ramesh Kumar Muralimanohar, and Kristin Sommerfeldt, "Externalization of remote microphone signals using a structural binaural model of the head and pinna," *The Journal of the Acoustical Society of America*, vol. 143, no. 5, pp. 2666–2677, 2018.

[20] Ryan M. Corey and Andrew C. Singer, "Adaptive binaural filtering for a multiple-talker listening system using remote and on-ear microphones," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2021.

[21] Harvey Dillon, *Hearing Aids*, Thieme, 2012.

[22] Jin Sook Kim and Chun Hyeok Kim, "A review of assistive listening device and digital wireless technology for hearing instruments," *Korean Journal of Audiology*, vol. 18, no. 3, pp. 105–111, 2014.

[23] Sergei Kochkin, Juliëtte Sterkens, Cynthia Compton-Conley, Douglas L. Beck, Brian Taylor, Patricia Kricos, Mary Caccavo, Alice Holmes, and Thomas A. Powers, "Consumer perceptions of the impact of inductively looped venues on the utility of their hearing devices," *The Hearing Review*, vol. 35, no. 5, pp. 16–26, 2014.

[24] Nick Hunn, *Introducing Bluetooth LE Audio*, Independently published, 2022.

[25] Simon Haykin, *Adaptive Filter Theory*, Prentice Hall, 2002.

[26] Eberhard Hänsler and Gerhard Schmidt, *Acoustic Echo and Noise Control: A Practical Approach*, Wiley, 2004.

[27] Christophe Veaux, Junichi Yamagishi, and Kirsten MacDonald, "CSTR VCTK corpus: English multi-speaker corpus for CSTR voice cloning toolkit," University of Edinburgh. The Centre for Speech Technology Research (CSTR), 2017.

[28] Shmulik Markovich-Golan, Sharon Gannot, and Israel Cohen, "Blind sampling rate offset estimation and compensation in wireless acoustic sensor networks with application to beamforming," in *International Workshop on Acoustic Signal Enhancement (IWAENC)*, 2012.