EFFECT OF TARGET SIGNALS AND DELAYS ON SPATIALLY SELECTIVE ACTIVE NOISE CONTROL FOR OPEN-FITTING HEARABLES

Tong Xiao, Simon Doclo

Department of Medical Physics and Acoustics and Cluster of Excellence Hearing4all,
Carl von Ossietzky Universität Oldenburg, Germany

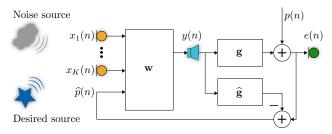tong.xiao@uni-oldenburg.de   simon.doclo@uni-oldenburg.de

ABSTRACT

Spatially selective active noise control (ANC) hearables are designed to reduce unwanted noise from certain directions while preserving desired sounds from other directions. In previous studies, the target signal has been defined either as the delayed desired component in one of the reference microphone signals or as the desired component in the error microphone signal without any delay. In this paper, we systematically investigate the influence of delays in different target signals on the ANC performance and provide an intuitive explanation for how the system obtains the desired signal. Simulations were conducted on a pair of open-fitting hearables for localized speech and noise sources in an anechoic environment. The performance was assessed in terms of noise reduction, signal quality and control effort. Results indicate that optimal performance is achieved without delays when the target signal is defined at the error microphone, whereas causality necessitates delays when the target signal is defined at the reference microphone. The optimal delay is found to be the acoustic delay between this reference microphone and the error microphone from the desired source.

Index Terms— Active noise control, spatial selectivity, beamforming, signal delay, control effort

1. INTRODUCTION

Active noise control (ANC) hearables are designed to create a quiet environment by using secondary sources to generate anti-noise, aiming at minimizing sound at certain positions when superimposed on the primary noise [1], [2]. Based on their fit, hearables can be categorized as closed-fitting (completely occluding the ear), open-fitting (partially occluding the ear), and open-ear (no occlusion). Recent research focuses on designing intelligent ANC hearables with spatial selectivity, especially for complex acoustic environments like cocktail-party scenarios with multiple sound sources from different directions [3]–[5]. In these environments, users may want to focus on desired sounds from a specific direction (e.g., from the front) while blocking out undesired sounds from other directions.

Modern ANC hearables are commonly equipped with multiple microphones, including both reference microphones on the exterior of the hearable and error microphones in the interior close to the ear canal. Hence, beamforming can be used to enhance a sound source from a certain direction and reduce sound sources from other directions [6]–[8], e.g., using the linearly-constrained-minimum-power (LCMP) or the minimum-power-distortionless-response (MPDR) beamformer. While traditional beamforming relies solely on the

Fig. 1. Block diagram of an ANC system with K reference microphones, one loudspeaker and one error microphone. The control filter is $\mathbf{w}$, the secondary path is denoted by $\mathbf{g}$ and its estimate is denoted by $\hat{\mathbf{g}}$.

microphone signals and thus performs passive noise reduction, recent advancements have proposed integrating a beamformer into an ANC system, i.e., performing noise reduction by jointly processing the microphone signals and playing back anti-noise through the loudspeakers [9]–[12]. These studies have considered two types of target signals, each associated with different delays. The first approach defines the target signal as the delayed desired component in one of the reference microphone signals (or a linear combination) [9]–[11], where the delay is typically chosen to be half of the filter length to maintain causality. The second approach defines the target signal as the desired component in the error microphone signal without any delay [12]. While the first approach is suitable for any fit, the second approach is preferable for open-ear and open-fitting hearables.

In this paper, we systematically investigate the influence of delays for different target signals on the performance of spatially selective ANC for open-fitting hearables. The range of considered delays spans from zero to half of the filter length, which is a common choice for passive noise reduction algorithms. The findings identify an optimal range of delays, particularly when considering the control effort. Moreover, the delay analysis enables us to provide an intuitive explanation for how the system obtains the target signal.

2. SIGNAL MODEL

As shown in Fig. 1, we consider an ANC system with K reference microphones. Without loss of generality, we consider one loudspeaker as the secondary source and one error microphone, resulting in a total of $K + 1$ microphones. We assume that acoustic feedback from the secondary source to the reference microphones can be perfectly canceled. We assume that the desired sound is speech and is coming from a different direction than the undesired noise. Subscripts $(\cdot)_s$ and $(\cdot)_v$ denote the speech and noise components in signals, respectively.

Before ANC is enabled, the primary sound signal $p(n)$ at the error microphone is given by,

$$p(n) = p_s(n) + p_v(n), \qquad (1)$$

where n denotes the time index. After ANC is enabled, the error signal is

$$e(n) = e_s(n) + e_v(n). \qquad (2)$$

The anti-noise at the error microphone can be represented by the product of the stacked input vector $\mathbf{x}(n)$ with the stacked control filter $\mathbf{w}$ and the convolution matrix $\widetilde{\mathbf{G}}$ of the secondary path, i.e.,

$$e(n) = p(n) + (\widetilde{\mathbf{G}}\mathbf{w})^{\mathrm{T}}\mathbf{x}(n), \qquad (3)$$

where superscript $(\cdot)^{\mathrm{T}}$ denotes the transpose. The stacked control filter $\mathbf{w}$ is defined as

$$\mathbf{w} = \begin{bmatrix} \mathbf{w}_1^{\mathrm{T}} \ \mathbf{w}_2^{\mathrm{T}} \ \dots \ \mathbf{w}_{K+1}^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}} \in \mathbb{R}^{(K+1)L_w}, \qquad (4a)$$

$$\mathbf{w}_k = [w_{k,0} \ w_{k,1} \ \dots \ w_{k,L_w-1}]^{\mathrm{T}} \in \mathbb{R}^{L_w}, \qquad (4b)$$

where L_w denotes the control filter length for each channel. The convolution matrix of the secondary path is given by

$$\widetilde{\mathbf{G}} = \mathrm{blkdiag}\,(\mathbf{G}\ \mathbf{G}\ \dots\ \mathbf{G})\,, \qquad (5a)$$

$$\mathbf{G} = \begin{bmatrix} g_0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ g_{L_g-1} & & g_0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & g_{L_g-1} \end{bmatrix} \in \mathbb{R}^{(L_g+L_w-1)\times L_w}, \qquad (5b)$$

where L_g is the secondary path filter length. The stacked input vector $\mathbf{x}(n)$ is defined as

$$\mathbf{x}(n) = \begin{bmatrix} \mathbf{x}_1^{\mathrm{T}}(n) \ \dots \ \mathbf{x}_K^{\mathrm{T}}(n) \ \widehat{\mathbf{p}}^{\mathrm{T}}(n) \end{bmatrix}^{\mathrm{T}} \in \mathbb{R}^{(K+1)(L_g+L_w-1)}, \quad (6)$$

with

$$\mathbf{x}_k(n) = [x_k(n) \ \dots \ x_k(n - L_g - L_w + 2)]^{\mathrm{T}}, \qquad (7a)$$

$$\widehat{\mathbf{p}}(n) = [\widehat{p}(n) \ \dots \ \widehat{p}(n - L_g - L_w + 2)]^{\mathrm{T}}, \qquad (7b)$$

where $\widehat{p}(n)$ is an estimate of the primary sound signal $p(n)$. This estimate can be computed from the error signal $e(n)$ and the secondary source signal $y(n)$ as

$$\widehat{p}(n) = e(n) - \widehat{\mathbf{g}}^{\mathrm{T}}\mathbf{y}(n), \qquad (8)$$

where $\widehat{\mathbf{g}}$ denotes an estimate of the secondary path $\mathbf{g} = [g_0 \ g_1 \ \dots \ g_{L_g-1}]^{\mathrm{T}}$.

Assuming a perfect estimate of the secondary path to be available, i.e., $\widehat{\mathbf{g}} = \mathbf{g}$, such that $\widehat{p}(n) = p(n)$, the primary sound signal can be written as $p(n) = \mathbf{q}^{\mathrm{T}}\mathbf{x}(n)$, with

$$\mathbf{q} = \begin{bmatrix} \mathbf{0}^{\mathrm{T}} \ \mathbf{0}^{\mathrm{T}} \ \dots \ \mathbf{0}^{\mathrm{T}} \ \boldsymbol{\delta}^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}} \in \mathbb{R}^{(K+1)(L_g+L_w-1)}, \qquad (9a)$$

$$\boldsymbol{\delta} = [1 \ 0 \ \dots \ 0]^{\mathrm{T}} \in \mathbb{R}^{(L_g+L_w-1)}, \qquad (9b)$$

such that the error signal in (3) can be written as

$$e(n) = \mathbf{q}^{\mathrm{T}}\mathbf{x}(n) + (\widetilde{\mathbf{G}}\mathbf{w})^{\mathrm{T}}\mathbf{x}(n) = (\mathbf{q} + \widetilde{\mathbf{G}}\mathbf{w})^{\mathrm{T}}\mathbf{x}(n). \qquad (10)$$

3. SPATIALLY SELECTIVE ANC

Conventional ANC systems minimize sounds regardless of their incoming directions. A spatially selective ANC system integrates a beamformer into the system such that only undesired sounds from certain directions are minimized, while the desired sound remains. The desired components in the input signals can be separated from the undesired components by computing the relative impulse responses (ReIRs) of the desired source.

3.1. Cost function and solution

We denote the target signal as $t(n)$, which is the signal the system aims to obtain at the error microphone after ANC. There are various possible choices of the target signal in a spatially selective ANC system. The target signal can be defined as a filtered version of the desired component in either one of the reference microphone signals or the error microphone signal. Similar to MPDR and LCMP beamformers, while also minimizing the power of the error signal $e(n)$, a spatial constraint based on the ReIRs of the desired source can be applied to the control filter, i.e.,

$$\mathbf{H}^{\mathrm{T}}(\mathbf{q} + \widetilde{\mathbf{G}}\mathbf{w}) = \mathbf{f}, \qquad (11)$$

with

$$\mathbf{H} = [\mathbf{H}_1 \ \mathbf{H}_2 \ \dots \ \mathbf{H}_{K+1}]^{\mathrm{T}} \in \mathbb{R}^{(K+1)L\times(L_h+L-1)}, \qquad (12)$$

where $\mathbf{H}_k \in \mathbb{R}^{(L_h+L-1)\times L}$ is the convolution matrix with a similar form as (5b) of the ReIR $\mathbf{h}_k = [h_{k,0} \ h_{k,1} \ \dots \ h_{k,L_h-1}]^{\mathrm{T}}$ between the k-th microphone and a chosen *spatial reference microphone* with L_h being the ReIR filter length and $L = L_g + L_w - 1$. All ReIRs can be determined from the acoustic impulse responses between the desired source and the microphones. Here, we take the microphone closest to the desired source as the *spatial reference microphone* and assume all ReIRs to be causal. The constraint vector $\mathbf{f} \in \mathbb{R}^{L_h+L-1}$ in (11) reflects the target signal $t(n)$. It can have different definitions depending on different choices of the target signal, which will be discussed in the next subsection.

Using (10) and (11), the cost function for a spatially selective ANC system can be defined as

$$\min_{\mathbf{w}} E\left\{e^2(n)\right\} + \beta\mathbf{w}^{\mathrm{T}}\mathbf{w}$$
$$= \min_{\mathbf{w}} E\left\{(\mathbf{q} + \widetilde{\mathbf{G}}\mathbf{w})^{\mathrm{T}}\Phi_{\mathbf{xx}}(\mathbf{q} + \widetilde{\mathbf{G}}\mathbf{w})\right\} + \beta\mathbf{w}^{\mathrm{T}}\mathbf{w}$$
$$\text{s. t. } \mathbf{H}^{\mathrm{T}}(\mathbf{q} + \widetilde{\mathbf{G}}\mathbf{w}) = \mathbf{f}, \qquad (13)$$

where $\Phi_{\mathbf{xx}} = E\left\{\mathbf{x}(n)\mathbf{x}^{\mathrm{T}}(n)\right\}$ is the autocorrelation matrix of the input vector, with $E\{\cdot\}$ being the mathematical expectation operator, and β being a control effort weighting factor.

The solution of (13) is given by [12]

$$\mathbf{w} = -\left[\mathbf{I} - \Phi_{\mathbf{rr}}^{-1}\widetilde{\mathbf{G}}^{\mathrm{T}}\mathbf{H}(\mathbf{H}^{\mathrm{T}}\widetilde{\mathbf{G}}\Phi_{\mathbf{rr}}^{-1}\widetilde{\mathbf{G}}^{\mathrm{T}}\mathbf{H} + \rho\mathbf{I})^{-1}\mathbf{H}^{\mathrm{T}}\widetilde{\mathbf{G}}\right]\Phi_{\mathbf{rr}}^{-1}\phi$$
$$+ \Phi_{\mathbf{rr}}^{-1}\widetilde{\mathbf{G}}^{\mathrm{T}}\mathbf{H}(\mathbf{H}^{\mathrm{T}}\widetilde{\mathbf{G}}\Phi_{\mathbf{rr}}^{-1}\widetilde{\mathbf{G}}^{\mathrm{T}}\mathbf{H} + \rho\mathbf{I})^{-1}\left(\mathbf{f} - \mathbf{H}^{\mathrm{T}}\mathbf{q}\right), \quad (14)$$

with

$$\Phi_{\mathbf{rr}} = \widetilde{\mathbf{G}}^{\mathrm{T}}\Phi_{\mathbf{xx}}\widetilde{\mathbf{G}} + \beta\mathbf{I}, \qquad \phi = \widetilde{\mathbf{G}}^{\mathrm{T}}\Phi_{\mathbf{xx}}\mathbf{q}, \qquad (15)$$

where $\mathbf{I}$ denotes the identity matrix, and ρ is a regularization factor due to matrix $\mathbf{H}^{\mathrm{T}}\widetilde{\mathbf{G}}\Phi_{\mathbf{rr}}^{-1}\widetilde{\mathbf{G}}^{\mathrm{T}}\mathbf{H}$ possibly being rank-deficient (e.g., due to delays in the secondary path) [13].

3.2. Target signal and delays

Studies [9]–[12] all similarly minimized the noise component. However, they had different target signals to obtain.

In [9]–[11], the target signal was defined as the delayed desired component at a reference microphone (e.g., the *spatial reference microphone*), that is, $t(n) = x_{\mathrm{ref},s}(n - \Delta)$. In this case, vector $\mathbf{f}$ is given by

$$\mathbf{f} = \boldsymbol{\Psi}\boldsymbol{\delta}_\Delta, \quad \boldsymbol{\delta}_\Delta = [\underbrace{0 \ \ldots \ 0}_{\Delta} \ 1 \ 0 \ \ldots \ 0]^{\mathrm{T}} \in \mathbb{R}^{L_h}, \quad (16)$$

where the spectral weighting matrix $\boldsymbol{\Psi} \in \mathbb{R}^{(L_h + L - 1) \times L_h}$ is the convolution matrix from a minimum-phase high-pass filter $\boldsymbol{\psi} \in \mathbb{R}^L$ (e.g., cut off at 120 Hz). Such a spectral weighting method can be used to improve the noise reduction performance at the cost of some signal distortion [12].

In [12], the target signal was defined as the desired component at the error microphone, which may also include certain delays, i.e., $t(n) = p_s(n - \Delta)$. In this case, vector $\mathbf{f}$ is given by

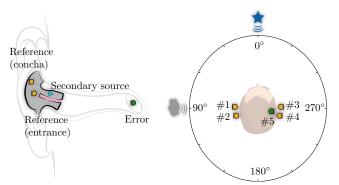$$\mathbf{f} = \boldsymbol{\Psi}\mathbf{h}_{K+1,\Delta}, \quad (17)$$

where $\mathbf{h}_{K+1,\Delta}$ is the ReIR from the spatial reference microphone to the error microphone $\mathbf{h}_{K+1}$ delayed by Δ samples.

4. SIMULATIONS

In this section, we systematically evaluate the influence of delays on the system performance. Section 4.1 discusses the acoustic setup, algorithm parameters and evaluation metrics. Sections 4.2 and 4.3 present simulation results when defining the target signal either at the error microphone or at the spatial reference microphone.

4.1. Setup and evaluation metrics

For the simulations, we considered a pair of open-fitting hearables [14], [15] inserted in both ears of a GRAS 45BB-12 KEMAR Head & Torso simulator, as shown in Fig. 2. We used four reference microphones (concha microphones at the left and right ears, entrance microphones at the left and right ears, labeled as #1 – #4), one error microphone (located at the right ear, labeled as #5) and one secondary source (outer receiver at the right ear). The error microphone was assumed to be at the eardrum. To generate the microphone signals and compute the ReIRs, we used the database from [15], which contains measured impulse responses in an anechoic chamber for a source at various directions relative to these microphones. In this setup, we considered a desired clean speech source from $0°$ ("p234_005" from the VCTK dataset [16]) and a noise source from $90°$ (babble noise from the NOISEX-92 database [17]). The signals had a duration of 5 s with a sampling rate of 16 kHz. The signal-to-noise ratio (SNR) at the error microphone was set to –5 dB. The filter lengths of the control filter, the secondary path and the ReIRs were equal to $L_w = L_g = L_h = 280$. The causal ReIRs were computed using the least-mean-squares adaptive filter (after convergence) from microphone signals simulated with white noise at $0°$, using the entrance microphone #3 as the spatial reference microphone. The high-pass filter $\boldsymbol{\psi}$ had a cut-off frequency at 120 Hz. In all cases, $\beta = \lambda_{\max}(\widetilde{\mathbf{G}}^{\mathrm{T}}\Phi_{\mathbf{xx}}\widetilde{\mathbf{G}})/500$ and $\rho = \lambda_{\max}(\mathbf{H}^{\mathrm{T}}\widetilde{\mathbf{G}}\Phi_{\mathbf{rr}}^{-1}\widetilde{\mathbf{G}}^{\mathrm{T}}\mathbf{H})/30000$, where $\lambda_{\max}(\cdot)$ denotes the largest eigenvalue.

Fig. 2. Illustration of the open-fitting hearable, the considered microphones and the simulation setup.

The following four metrics are used for evaluation. The noise reduction (NR) level of the ANC is defined as

$$\mathrm{NR \ (dB)} = 10 \log_{10} \left(\frac{\sum\limits_{n=1}^{N} p_v^2(n)}{\sum\limits_{n=1}^{N} e_v^2(n)} \right), \quad (18)$$

where $e_v(n)$ can be obtained by using $\mathbf{w}$ in (14) to filter only the noise components in the signals. N denotes the signal length.

The speech distortion index (SDI) is used to assess distortion in the speech component at the error microphone after ANC [18]. It is defined as

$$\mathrm{SDI \ (dB)} = 10 \log_{10} \left(\frac{\sum\limits_{n=1}^{N} [t(n) - e_s(n)]^2}{\sum\limits_{n=1}^{N} t^2(n)} \right), \quad (19)$$

where $e_s(n)$ can be obtained by using $\mathbf{w}$ to filter only the speech components in the signals.

To assess the overall signal quality, we used the narrowband perceptual evaluation of speech quality (NB-PESQ) mean opinion score - listening quality objective (MOS-LQO) [19], either between $t(n)$ and $p(n)$ (ANC off) or between $t(n)$ and $e(n)$ (ANC on). It should be noted that the reference signals in the two cases are different, since the definitions of the target signal are different.
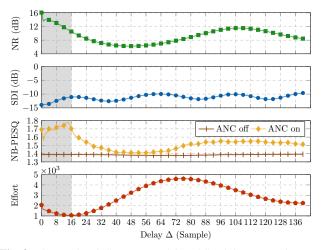
Finally, we computed the control effort, an important factor to consider in ANC systems [1], which is defined as

$$\mathcal{E} = \sum_{n=1}^{N} y^2(n). \quad (20)$$

4.2. Target signal at the error microphone

Figure 3 depicts the evaluation metrics for various delays, ranging from 0 to 140 (half of the control filter length, $L_w/2$) with a step size of 1, when the target signal is defined as the delayed speech component at the error microphone, i.e., $t(n) = p_s(n - \Delta)$.

For all considered metrics, it can be observed that the best performance is obtained for small delays. In fact, for all metrics except control effort, the best performance is obtained for $\Delta = 0$. The lowest value for the control effort is obtained for $\Delta = 16$, but for this value, the NR level is also degraded. For example, the NR level is 16.1 dB for $\Delta = 0$, but 10.5 dB for $\Delta = 16$, although the control effort decreased from 2065 to 1063.

Fig. 3. The NR level, SDI, NB-PESQ MOS-LQO score and control effort for $\Delta \in [0 : 1 : 140]$ sample delays in the target signal (desired speech component at the *error* microphone). The shaded areas indicate the recommended range for the delay.
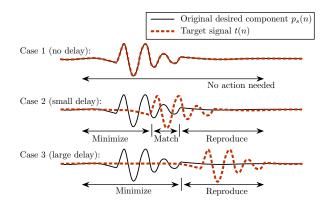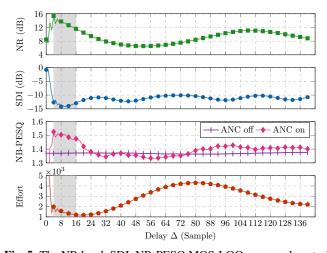

Fig. 4. Illustration of the mechanism behind selective ANC for various delays between the original desired component at the error microphone $p_s(n)$ and the target signal $t(n)$.

For three cases (no delay, small delay, large delay), Fig. 4 visualizes the mechanism behind selective ANC, depicting the original speech component in the error microphone, $p_s(n)$, and the target signal, $t(n) = p_s(n - \Delta)$. For Case 1 (no delay), the original speech component just needs to be preserved, requiring no action for this system [12]. For Case 2 (small delay), there are three mechanisms involved: minimizing the original speech component, matching the original speech component to the target, and reproducing the target signal. This will typically require more control effort than for $\Delta = 0$. The extreme case is shown in Case 3 (large delay), where the original speech component and the target are completely misaligned due to the delay. The system will need to minimize the original speech component first and then reproduce the target once again, thus requiring the most effort. This may explain why for $\Delta \geq 40$, there is a significant increase of the control effort but with low NR levels.

In summary, these simulation results indicate that although a slight delay is permissible ($\Delta \leq 16$, as seen in the shaded areas in Fig. 3), it is still best to design the delay equal to zero when the target is defined at the error microphone.

Fig. 5. The NR level, SDI, NB-PESQ MOS-LQO score and control effort for $\Delta \in [0 : 1 : 140]$ sample delays in the target signal (desired speech component at the *spatial reference* microphone). The shaded areas indicate the recommended range for the delay.

4.3. Target signal at the spatial reference microphone

Figure 5 depicts the evaluation metrics for the same various delays, when the target signal is defined as the delayed speech component in the spatial reference microphone, i.e., $t(n) = x_{\mathrm{ref},s}(n - \Delta)$. For all considered metrics, it can be observed that the worst performance is obtained for $\Delta = 0$. Since the system tries to obtain the original speech component from the spatial reference microphone at the error microphone, causality cannot be satisfied [9], [20], [21], thus requiring certain delays.

In [9], [10], a delay of half of the filter length was used. For our simulations, using $\Delta = 140$ results in 8.9 dB NR, -10.8 dB SDI, 0.03 PESQ improvement and a control effort of 2207. However, as depicted in the shaded areas of Fig. 5, it is preferable to apply only a minor delay ($4 \leq \Delta \leq 16$) to achieve the largest NR and PESQ improvement, as well as the lowest distortion and control effort. For example, $\Delta = 4$ results in 15.4 dB NR level, -12.7 dB SDI, 0.16 PESQ improvement with a control effort of 1983. In fact, the acoustic delay between the spatial reference microphone and the error microphone from the desired source also has four samples. Therefore, it is sensible to allow for a four-sample delay in the system.

In summary, these simulation results indicate that when the target signal is defined in a reference microphone, excessive delays result in inefficient control and small (non-zero) delays are preferred instead. The mechanism can be similarly explained as in Fig. 4.

5. CONCLUSION

This paper examined two types of target signals with various delays in a pair of spatially selective open-fitting ANC hearables. When the target signal is the desired component in the error microphone signal, optimal performance was achieved when the original desired component was preserved without any delay. However, when the target signal is the desired component in a reference microphone signal, a small delay is required to satisfy causality. The optimal delay is found to be the acoustic delay between this reference microphone and the error microphone from the desired source. Using large delays leads to degraded performance and increased control effort as the system attempts to minimize the original speech component before reproducing the delayed version.

6. REFERENCES

[1] S. J. Elliott, *Signal processing for active control*. Academic Press, 2000.

[2] C. Hansen, S. Snyder, X. Qiu, L. Brooks, and D. Moreau, *Active Control of Noise and Vibration*, 2nd ed. CRC Press, Nov. 2012.

[3] Y. Kajikawa, W.-S. Gan, and S. M. Kuo, "Recent advances on active noise control: Open issues and innovative applications," *APSIPA Transactions on Signal and Information Processing*, vol. 1, e3, 2012.

[4] C.-Y. Chang, A. Siswanto, C.-Y. Ho, T.-K. Yeh, Y.-R. Chen, and S. M. Kuo, "Listening in a noisy environment: Integration of active noise control in audio products," *IEEE Consumer Electronics Magazine*, vol. 5, no. 4, pp. 34–43, 2016.

[5] R. Gupta, J. He, R. Ranjan, W.-S. Gan, F. Klein, C. Schneiderwind, A. Neidhardt, K. Brandenburg, and V. Välimäki, "Augmented/mixed reality audio for hearables: Sensing, control, and rendering," *IEEE Signal Processing Magazine*, vol. 39, no. 3, pp. 63–89, 2022.

[6] B. Van Veen and K. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE ASSP Magazine*, vol. 5, no. 2, pp. 4–24, 1988.

[7] S. Doclo, W. Kellermann, S. Makino, and S. E. Nordholm, "Multichannel signal enhancement algorithms for assisted listening devices: Exploiting spatial diversity using multiple microphones," *IEEE Signal Processing Magazine*, vol. 32, no. 2, pp. 18–30, Mar. 2015.

[8] S. Gannot, E. Vincent, S. Markovich-Golan, and A. Ozerov, "A consolidated perspective on multimicrophone speech enhancement and source separation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 4, pp. 692–730, 2017.

[9] R. Serizel, M. Moonen, J. Wouters, and S. H. Jensen, "Integrated active noise control and noise reduction in hearing aids," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 6, pp. 1137–1146, 2010.

[10] D. Dalga and S. Doclo, "Combined feedforward-feedback noise reduction schemes for open-fitting hearing aids," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, NY, USA, 2011, pp. 185–188.

[11] V. Patel, J. Cheer, and S. Fontana, "Design and Implementation of an Active Noise Control Headphone with Directional Hear-Through Capability," *IEEE Transactions on Consumer Electronics*, vol. 66, no. 1, pp. 32–40, Feb. 2020.

[12] T. Xiao, B. Xu, and C. Zhao, "Spatially selective active noise control systems," *The Journal of the Acoustical Society of America*, vol. 153, no. 5, pp. 2733–2744, May 2023.

[13] P. C. Hansen, *Rank-deficient and discrete ill-posed problems: numerical aspects of linear inversion*. SIAM, 1998.

[14] F. Denk, M. Lettau, H. Schepker, S. Doclo, R. Roden, M. Blau, J.-H. Bach, J. Wellmann, and B. Kollmeier, "A one-size-fits-all earpiece with multiple microphones and drivers for hearing device research," in *Proc. AES International Conference on Headphone Technology*, San Francisco, USA, Aug. 2019.

[15] F. Denk and B. Kollmeier, "The hearpiece database of individual transfer functions of an in-the-ear earpiece for hearing device research," *Acta Acustica*, vol. 5, p. 2, 2021.

[16] C. Veaux, J. Yamagishi, K. MacDonald, *et al.*, "CSTR VCTK corpus: English multi-speaker corpus for CSTR voice cloning toolkit," *University of Edinburgh. The Centre for Speech Technology Research (CSTR)*, vol. 6, p. 15, 2017.

[17] A. Varga and H. J. Steeneken, "Assessment for automatic speech recognition: II. NOISEX-92: A database and an experiment to study the effect of additive noise on speech recognition systems," *Speech Communication*, vol. 12, no. 3, pp. 247–251, 1993.

[18] J. Chen, J. Benesty, Y. Huang, and S. Doclo, "New insights into the noise reduction Wiener filter," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 4, pp. 1218–1234, 2006.

[19] A. Rix, J. Beerends, M. Hollier, and A. Hekstra, "Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (ICASSP)*, Salt Lake City, Utah, USA, May 2001, pp. 749–752.

[20] R. Ranjan and W.-S. Gan, "Natural listening over headphones in augmented reality using adaptive filtering techniques," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 11, pp. 1988–2002, 2015.

[21] H. Schepker, F. Denk, B. Kollmeier, and S. Doclo, "Robust single-and multi-loudspeaker least-squares-based equalization for hearing devices," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2022, no. 1, pp. 1–14, 2022.