

# Soft-Constrained Spatially Selective Active Noise Control for Open-fitting Hearables

Tong Xiao<sup>1</sup>, Reinhild Roden<sup>2</sup>, Matthias Blau<sup>2</sup>, Simon Doclo<sup>1</sup>

<sup>1</sup> 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

<sup>2</sup> Institut für Hörtechnik und Audiologie and Cluster of Excellence Hearing4all, Jade-Hochschule, Oldenburg, Germany

reinhild.roden@jade-hs.de, matthias.blau@jade-hs.de

**Abstract**—Recent advances in spatially selective active noise control (SSANC) using multiple microphones have enabled hearables to suppress undesired noise while preserving desired speech from a specific direction. Aiming to achieve minimal speech distortion, a hard constraint has been used in previous work in the optimization problem to compute the control filter. In this work, we propose a soft-constrained SSANC system that uses a frequency-independent parameter to trade off between speech distortion and noise reduction. We derive both time- and frequency-domain formulations, and show that conventional active noise control and hard-constrained SSANC represent two limiting cases of the proposed design. We evaluate the system through simulations using a pair of open-fitting hearables in an anechoic environment with one speech source and two noise sources. The simulation results validate the theoretical derivations and demonstrate that for a broad range of the trade-off parameter, the signal-to-noise ratio and the speech quality and intelligibility in terms of PESQ and ESTOI can be substantially improved compared to the hard-constrained design.

## 1. INTRODUCTION

Active noise control (ANC) hearables aim to create a quiet environment by using secondary sources to generate anti-noise, minimizing sound when superimposed on the primary noise (also referred to as leakage) [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; e.g., smart glasses). While open-fitting and open-ear designs reduce the occlusion effect and improve physical comfort, they also lead to increased leakage. Recent research focuses on intelligent ANC hearables with spatial selectivity for complex acoustic environments involving multiple sound sources from different directions [3]–[8]. In such scenarios, listeners often want to focus on sound from a particular direction (e.g., speech from the front) while suppressing noise leakage from other directions.

Modern ANC hearables are commonly equipped with multiple microphones, i.e., microphones on the exterior of the hearable and error microphones in the interior close to the eardrum. On the one hand, conventional ANC algorithms suppress all leakage measured by the inner error microphones, including desired speech [1], [2], [9]. On the other hand, multi-microphone noise reduction algorithms, e.g., minimum variance distortionless response (MVDR) beamforming [10]–[12], perform spatial filtering by reducing all noise while preserving speech from the desired direction. However, these algorithms typically ignore the leakage and do not exploit the inner error microphones. Recent work on spatially selective active noise control (SSANC) integrates spatial filtering into ANC systems, thus preserving speech from a desired direction while actively suppressing sound from undesired directions [6]–[8], [13]–[21]. To achieve minimal speech distortion, the SSANC algorithm in [8] imposes a hard constraint

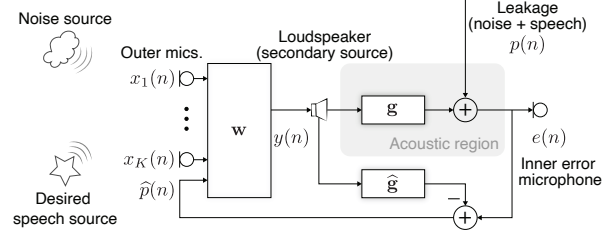


Fig. 1: Block diagram of an ANC system with  $K$  outer microphones, one inner error microphone and one loudspeaker (i.e., secondary source). The control filter is denoted by  $w$ , the secondary path is denoted by  $g$ , and its estimate by  $\hat{g}$ .

in the optimization problem for the control filter. For this algorithm, it has been shown in [20] that when aiming to preserve the speech component in an outer reference microphone signal, a delay equal to the acoustic propagation delay between the reference microphone and the inner error microphone for the desired source is necessary to satisfy causality. Furthermore, it has been shown in [21] that using acausal relative impulse responses in the filter optimization is beneficial.

In this paper, we introduce a soft-constrained SSANC system that employs a single frequency-independent parameter to adjust the trade-off between speech distortion and noise reduction. Although soft-constrained design procedures have been explored for multi-microphone speech enhancement [22]–[26], e.g., speech distortion weighted multichannel Wiener filter (SDW-MWF), and for feedback suppression [27], they have rarely been studied for ANC. We derive both time- and frequency-domain formulations for the conventional ANC, the hard-constrained SSANC, and the proposed soft-constrained SSANC. For the frequency-domain formulations, we show that conventional ANC and hard-constrained SSANC correspond to limiting cases of the soft-constrained design, validating these theoretical relationships through causal time-domain simulations. For the considered setup, simulation results demonstrate that for a broad range of the trade-off parameter, the proposed soft-constrained SSANC outperforms both hard-constrained SSANC and conventional ANC.

## 2. SIGNAL MODEL

As shown in Fig. 1, we consider a hearable with  $K$  outer microphones. Without loss of generality, we consider one loudspeaker as the secondary source and one inner error microphone, resulting in a total of  $K + 1$  microphones. We assume that the acoustic feedback paths between the loudspeaker and the outer microphones are known, such that acoustic feedback can be canceled. In addition, we assume that an estimate of the secondary path between the loudspeaker and the inner error microphone is available.

This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 352015383 – SFB 1330 C1.

The inner error microphone signal  $e(n)$ , with  $n$  the discrete time index, is given by

$$e(n) = p(n) + (\mathbf{G}\mathbf{w})^T \mathbf{x}(n), \quad (1)$$

where  $(\cdot)^T$  denotes the transpose operator. The leakage (including noise and desired speech) at the inner error microphone is denoted by  $p(n)$ , and the anti-noise component at the inner error microphone is given by  $(\mathbf{G}\mathbf{w})^T \mathbf{x}(n)$ , where  $\mathbf{w}$  is the stacked control filter,  $\mathbf{x}(n)$  is the stacked input vector, and  $\mathbf{G}$  represents the secondary path convolution matrix. The stacked control filter  $\mathbf{w}$  is defined as

$$\mathbf{w} = [\mathbf{w}_1^T \ \mathbf{w}_2^T \ \dots \ \mathbf{w}_{K+1}^T]^T \in \mathbb{R}^{(K+1)L_w}, \quad (2a)$$

$$\mathbf{w}_k = [w_{k,0} \ w_{k,1} \ \dots \ w_{k,L_w-1}]^T \in \mathbb{R}^{L_w}, \quad (2b)$$

where  $L_w$  denotes the control filter length for each channel. The convolution matrix of the secondary path is defined as

$$\mathbf{G} = \text{blkdiag}(\mathbb{G} \dots \mathbb{G}) \in \mathbb{R}^{(K+1)L \times (K+1)L_w}, \quad (3a)$$

$$\mathbb{G} = \begin{bmatrix} g_0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ g_{L_g-1} & \dots & g_0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & g_{L_g-1} \end{bmatrix} \in \mathbb{R}^{L \times L_w}, \quad (3b)$$

where  $L = L_g + L_w - 1$ , and  $L_g$  denotes the secondary path filter length. As input signals to the control filter we consider the  $K$  outer microphone signals  $\mathbf{x}_k(n)$ ,  $k = 1, \dots, K$ , and an estimate of the leakage  $\hat{\mathbf{p}}(n)$ , i.e., the stacked input vector  $\mathbf{x}(n)$  is defined as

$$\mathbf{x}(n) = [\mathbf{x}_1^T(n) \ \dots \ \mathbf{x}_K^T(n) \ \hat{\mathbf{p}}^T(n)]^T \in \mathbb{R}^{(K+1)L}, \quad (4)$$

with

$$\mathbf{x}_k(n) = [x_k(n) \ \dots \ x_k(n - L + 1)]^T, \quad (5a)$$

$$\hat{\mathbf{p}}(n) = [\hat{p}(n) \ \dots \ \hat{p}(n - L + 1)]^T. \quad (5b)$$

The estimated leakage  $\hat{p}(n)$  can be computed from the inner error microphone signal  $e(n)$ , the loudspeaker signal vector  $\mathbf{y}(n) = [y(n) \ \dots \ y(n - L_g + 1)]^T$ , and an estimate of the secondary path  $\hat{\mathbf{g}}$  as

$$\hat{p}(n) = e(n) - \hat{\mathbf{g}}^T \mathbf{y}(n). \quad (6)$$

Assuming a perfect estimate of the secondary path to be available, i.e.,  $\hat{\mathbf{g}} = \mathbf{g} = [g_0 \ g_1 \ \dots \ g_{L_g-1}]^T$ , the leakage can be written as  $p(n) = \hat{p}(n) = \mathbf{q}^T \mathbf{x}(n)$ , with

$$\mathbf{q} = [\mathbf{0}^T \ \dots \ \mathbf{0}^T \ \delta^T]^T \in \mathbb{R}^{(K+1)L}, \quad (7a)$$

$$\delta = [1 \ 0 \ \dots \ 0]^T \in \mathbb{R}^L. \quad (7b)$$

Hence, the inner error microphone signal in (1) can be rewritten as

$$e(n) = (\mathbf{q} + \mathbf{G}\mathbf{w})^T \mathbf{x}(n). \quad (8)$$

The frequency-domain representations of  $\mathbf{x}(n)$  and  $\mathbf{w}$  are

$$\mathbf{x}(\omega) = [X_1(\omega) \ \dots \ X_K(\omega) \ \hat{P}(\omega)]^T \in \mathbb{C}^{K+1}, \quad (9a)$$

$$\mathbf{w}(\omega) = [W_1(\omega) \ \dots \ W_K(\omega) W_{K+1}(\omega)]^T \in \mathbb{C}^{K+1}, \quad (9b)$$

where  $\omega$  denotes the radian frequency. Similar to (8), the frequency-domain inner error microphone signal is

$$E(\omega) = [\mathbf{q}_\omega^H + G(\omega)\mathbf{w}^H(\omega)]^T \mathbf{x}(\omega), \quad (10)$$

where  $(\cdot)^H$  denotes the Hermitian transpose, and  $\mathbf{q}_\omega = [0 \ \dots \ 0 \ 1]^T \in \mathbb{R}^{K+1}$ .  $G(\omega)$  denotes the transfer function of the secondary path.

### 3. CONVENTIONAL ANC AND HARD-CONSTRAINED SSANC IN TIME DOMAIN

The objective of conventional ANC is to minimize the power of the error microphone signal [1], [2], i.e.,

$$\min_{\mathbf{w}} \mathcal{E}\{e^2(n)\} + \beta \|\mathbf{w}\|_2^2, \quad (11)$$

where  $\mathcal{E}\{\cdot\}$  denotes the mathematical expectation operator.  $\beta$  is a regularization term.  $\|\cdot\|_2$  denotes the  $L_2$ -norm. Using (8), the solution of (11) is found to be

$$\mathbf{w}_{\text{ANC}} = -\Phi_{\text{rr}}^{-1} \Phi, \quad (12)$$

where

$$\Phi_{\text{rr}} = \mathbf{G}^T \Phi_{\text{xx}} \mathbf{G} + \beta \mathbf{I}, \quad (13a)$$

$$\Phi_{\text{xx}} = \mathcal{E}\{\mathbf{x}(n)\mathbf{x}^T(n)\}, \quad (13b)$$

$$\Phi = \mathbf{G}^T \Phi_{\text{xx}} \mathbf{q}, \quad (13c)$$

where  $\mathbf{I}$  denotes the identity matrix. It should be noted that both desired speech and noise are minimized in conventional ANC.

In [8], [20], [21], an SSANC system has been proposed where the objective is to minimize the power of the inner error microphone signal while preserving the delayed desired speech component of an outer reference microphone signal. This can be achieved by imposing the constraint

$$e_s(n) = x_{\text{ref},s}(n - \Delta), \quad (14)$$

where  $(\cdot)_s$  denotes the speech component of a signal and  $\Delta$  represents a delay. This constraint can be reformulated as [21]

$$\mathbf{H}(\mathbf{q} + \mathbf{G}\mathbf{w}) = \delta_\Delta, \quad (15)$$

where

$$\mathbf{H} = [\mathbf{H}_1 \ \dots \ \mathbf{H}_{K+1}] \in \mathbb{R}^{(L_a + L_h + L - 1) \times (K+1)L}, \quad (16a)$$

$$\mathbf{H}_k = \begin{bmatrix} h_{k,-L_a} & \dots & 0 \\ \vdots & \ddots & \vdots \\ h_{k,L_h-1} & \dots & 0 \\ 0 & \dots & h_{k,-L_a} \\ \vdots & \ddots & \vdots \\ 0 & \dots & h_{k,L_h-1} \end{bmatrix} \in \mathbb{R}^{(L_a + L_h + L - 1) \times L}, \quad (16b)$$

$$\delta_\Delta = [\underbrace{0 \ \dots \ 0}_{L_a} \ \underbrace{0 \ \dots \ 0}_\Delta \ \underbrace{1 \ 0 \ \dots \ 0}_{L_h + L - 1 - \Delta}]^T \in \mathbb{R}^{L_a + L_h + L - 1}, \quad (16c)$$

$\mathbf{H}_k$  is the convolution matrix of the relative impulse response (ReIR) between the  $k$ -th microphone and the reference microphone, with  $L_a$  and  $L_h$  denoting the length of the anti-causal and causal parts of the ReIR, respectively [21]. It should be particularly noted that although the ReIRs are acausal, the control filter  $\mathbf{w}$  is still causal.

To satisfy the distortionless constraint in (15), a hard constraint is added to the optimization problem in (11), i.e.,

$$\min_{\mathbf{w}} \mathcal{E}\{e^2(n)\} + \beta \|\mathbf{w}\|_2^2 \quad (17a)$$

$$\text{subject to } \mathbf{H}(\mathbf{q} + \mathbf{G}\mathbf{w}) = \delta_\Delta. \quad (17b)$$

The solution is found to be [8], [20], [21]

$$\mathbf{w}_{\text{hard}} = -\Phi_{\text{rr}}^{-1} \Phi + \mathbf{A} (\delta_\Delta - \mathbf{H}\mathbf{q} + \mathbf{H}\mathbf{G}\Phi_{\text{rr}}^{-1} \Phi), \quad (18)$$

with

$$\mathbf{A} = \Phi_{\text{rr}}^{-1} \mathbf{G}^T \mathbf{H}^T (\mathbf{H}\mathbf{G}\Phi_{\text{rr}}^{-1} \mathbf{G}^T \mathbf{H}^T + \rho \mathbf{I})^{-1}, \quad (19)$$

where a small regularization factor  $\rho$  is included to ensure the numerical stability of the matrix inversion when computing  $\mathbf{A}$ .

#### 4. SOFT-CONSTRAINED SSANC IN TIME DOMAIN

Since the optimization problem in (17) aims to perfectly preserve the speech component of the reference microphone signal at the inner error microphone, the noise reduction may be limited. By allowing a small amount of speech distortion, i.e., relaxing the distortionless constraint, a larger amount of noise reduction may be achieved, thus potentially improving signal-to-noise ratio and perceived speech quality and intelligibility.

To allow a trade-off between speech distortion and noise reduction, we propose a soft-constrained SSANC, by defining the optimization problem as

$$\min_{\mathbf{w}} \mathcal{E}\{e^2(n)\} + \beta \|\mathbf{w}\|_2^2 + \mu \|\mathbf{H}(\mathbf{q} + \mathbf{G}\mathbf{w}) - \delta_\Delta\|_2^2, \quad (20)$$

where  $\mu$  is a real-valued non-negative trade-off parameter. The solution is found to be

$$\mathbf{w}_{\text{soft}} = -\left(\Phi_{\text{rr}} + \mu \mathbf{G}^T \mathbf{H}^T \mathbf{H} \mathbf{G}\right)^{-1} \left[\Phi - \mu \mathbf{G}^T \mathbf{H}^T (\delta_\Delta - \mathbf{H}\mathbf{q})\right]. \quad (21)$$

When  $\mu = 0$ , the optimization problem in (20) becomes equal to (11), corresponding to the conventional ANC. Noise reduction is maximized in this case, but the desired speech component is also strongly suppressed (see simulations in Section 6). Increasing  $\mu$  gives more weight to the distortion term in (20), reducing speech distortion but potentially limiting the amount of noise reduction.

#### 5. FREQUENCY-DOMAIN ANALYSIS

In this section, we present a frequency-domain analysis for the conventional ANC, the hard-constrained SSANC, and the proposed soft-constrained SSANC, further highlighting their relationships.

Similarly to (11), the optimization problem for conventional ANC in the frequency domain is defined as

$$\min_{\mathbf{w}(\omega)} \mathcal{E}\{|E(\omega)|^2\}, \quad (22)$$

where the regularization term is omitted for brevity (i.e.,  $\beta = 0$ ). The solution can be found to be

$$\mathbf{w}_{\text{ANC}}(\omega) = \frac{1}{G^*(\omega)} (-\mathbf{q}\omega), \quad (23)$$

where  $(\cdot)^*$  denotes the complex conjugate.

Similarly to (17), the optimization problem for hard-constrained SSANC in the frequency domain is defined as

$$\min_{\mathbf{w}(\omega)} \mathcal{E}\{|E(\omega)|^2\} \quad (24a)$$

$$\text{subject to} \quad \left[\mathbf{q}_\omega^H + G(\omega)\mathbf{w}^H(\omega)\right] \mathbf{h}(\omega) = \exp(-i\omega\Delta), \quad (24b)$$

where  $\mathbf{h}(\omega) = [H_1(\omega) \dots H_K(\omega) H_{K+1}(\omega)]^T \in \mathbb{C}^{K+1}$  denotes the relative transfer function vector of the desired source, and  $i$  denotes the imaginary unit. The solution is found to be

$$\mathbf{w}_{\text{hard}}(\omega) = \frac{1}{G^*(\omega)} \left(-\mathbf{q}\omega + \frac{\Phi_x^{-1}(\omega) \mathbf{h}(\omega) \exp(i\omega\Delta)}{\mathbf{h}^H(\omega) \Phi_x^{-1}(\omega) \mathbf{h}(\omega)}\right), \quad (25)$$

where  $\Phi_x(\omega) = \mathcal{E}\{\mathbf{x}(\omega)\mathbf{x}^H(\omega)\} \in \mathbb{C}^{(K+1) \times (K+1)}$  denotes the input covariance matrix, which is assumed to be positive-definite.

Similarly to (20), the optimization problem for the proposed soft-constrained SSANC in the frequency domain is defined as

$$\min_{\mathbf{w}(\omega)} \mathcal{E}\{|E(\omega)|^2\} + \mu \left| \left[\mathbf{q}_\omega^H + G(\omega)\mathbf{w}^H(\omega)\right] \mathbf{h}(\omega) - \exp(-i\omega\Delta) \right|^2. \quad (26)$$

The solution is found to be

$$\mathbf{w}_{\text{soft}}(\omega) = \frac{1}{G^*(\omega)} \left(-\mathbf{q}\omega + \frac{\Phi_x^{-1}(\omega) \mathbf{h}(\omega) \exp(i\omega\Delta)}{\frac{1}{\mu} + \mathbf{h}^H \Phi_x^{-1}(\omega) \mathbf{h}(\omega)}\right). \quad (27)$$

By comparing (23), (25) and (27), it can be observed that

$$\lim_{\mu \rightarrow 0} \mathbf{w}_{\text{soft}}(\omega) = \mathbf{w}_{\text{ANC}}(\omega), \quad (28a)$$

$$\lim_{\mu \rightarrow \infty} \mathbf{w}_{\text{soft}}(\omega) = \mathbf{w}_{\text{hard}}(\omega), \quad (28b)$$

implying that in the frequency domain the conventional ANC and the hard-constrained SSANC are limiting cases of the soft-constrained SSANC for  $\mu \rightarrow 0$  and  $\mu \rightarrow \infty$ , respectively.

### 6. SIMULATION RESULTS

#### 6.1. Setup

For the simulation, we considered a pair of open-fitting hearables [28], [29] inserted into both ears of a GRAS 45BB-12 KEMAR Head & Torso simulator, as shown in Fig. 2a. We used four outer microphones (entrance microphones and concha microphones at the left and right ears, labeled as #1–#4 in Fig. 2b), one inner error microphone (located at the right ear, labeled as #5 in Fig. 2b), and the outer receiver at the right ear as the secondary source. The inner error microphone was assumed to be at the eardrum.

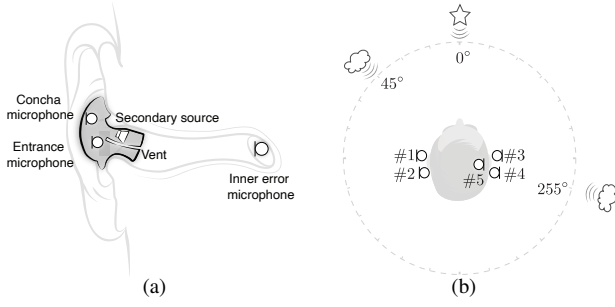
The acoustic scenario is shown in Fig. 2b, where we considered a desired speech source at  $0^\circ$  (“p361\_005” from the VCTK dataset [30]) and two noise sources at  $45^\circ$  and  $255^\circ$  (babble noise from the NOISEX-92 database [31]). The desired speech and noise components in all microphone signals were generated by convolving source signals with measured anechoic impulse responses from the database [29]. All signals were 5 seconds in duration and sampled at 16 kHz. The desired speech and noise components were mixed such that the signal-to-noise ratio (SNR) of the leakage at the inner error microphone was set to  $-5$  dB, with both noise sources contributing the same energy. As the secondary path estimate we used the measured impulse response between the outer receiver and the inner error microphone at the right ear (from [29]).

All the algorithms were implemented in the time domain. As filter lengths we used  $L_w = 600$  for the control filter (per channel),  $L_g = 280$  for the secondary path, and  $L_a = 22$  and  $L_h = 262$  for the anti-causal and causal parts of the ReIRs for the desired speech source in (16b). The delay for the desired speech component in (14) was set to  $\Delta = 32$ , corresponding to 2 ms. The ReIRs for the desired speech source were computed using the least-mean-squares adaptive filter (after convergence) from microphone signals generated by convolving white noise with the measured impulse responses at  $0^\circ$ , using the entrance microphone #3 as the reference microphone. Similar to the previous studies [8], [20], [21], the evaluation focused on the frequency range above 100 Hz. Therefore, a minimum-phase high-pass filter with a cut-off frequency at 100 Hz was applied to the desired signal  $x_{\text{ref},s}(n - \Delta)$ . For all ANC designs,  $\beta = \lambda_{\max}(\mathbf{G}^T \Phi_{\text{xx}} \mathbf{G}) / (4 \times 10^5)$ , where  $\lambda_{\max}(\cdot)$  denotes the largest eigenvalue. The regularization factor in (19) was equal to  $\rho = \lambda_{\max}(\mathbf{H} \mathbf{G} \Phi_{\text{rr}}^{-1} \mathbf{G}^T \mathbf{H}^T) / (4 \times 10^5)$ .

#### 6.2. Evaluation metrics

The performance of the considered ANC designs was evaluated in terms of noise reduction, speech distortion, SNR improvement, speech quality and intelligibility.

The noise reduction is defined as the difference between the power of the noise component of the leakage  $p_v(n)$  (without control) and



**Fig. 2:** (a) Illustration of the open-fitting hearable. (b) Acoustic scenario with one desired speech source at  $0^\circ$  and two babble noise sources at  $45^\circ$  and  $255^\circ$ .

the noise component of the inner error microphone signal  $e_v(n)$  (with control), i.e.,

$$\text{NR (dB)} = 10 \log_{10} \sum_{n=1}^N p_v^2(n) - 10 \log_{10} \sum_{n=1}^N e_v^2(n), \quad (29)$$

where  $N$  denotes the total signal length.

The intelligibility-weighted spectral distortion is used to assess the amount of speech distortion [6], [23], [25]. It is defined as

$$\text{SD}_{\text{intellig}} \text{ (dB)} = \sum_{b=1}^B I(\omega_b) 10 \log_{10} \frac{\mathcal{P}_\epsilon(\omega_b)}{\mathcal{P}_{\text{ref},s}(\omega_b)}, \quad (30)$$

where the band importance function  $I(\omega_b)$  expresses the importance of the  $b$ -th one-third octave band for intelligibility [32], and  $B$  denotes the total number of bands.  $\mathcal{P}_\epsilon(\omega_b)$  is the power spectral density of  $\epsilon(n)$  in the  $b$ -th band, where  $\epsilon(n) = e_s(n) - x_{\text{ref},s}(n - \Delta)$ .  $\mathcal{P}_{\text{ref},s}(\omega_b)$  is the power spectral density of  $x_{\text{ref},s}(n - \Delta)$  in the  $b$ -th band.

Similar to (30), the intelligibility-weighted SNR improvement is defined as

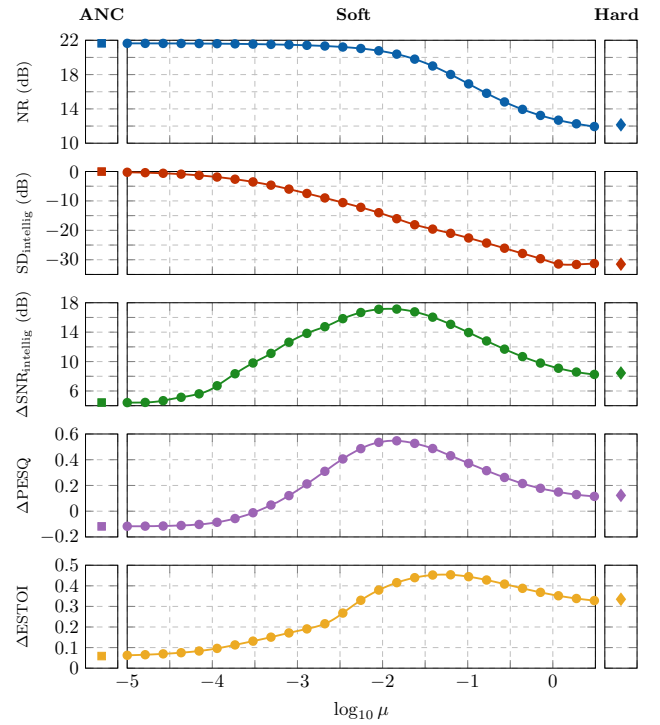
$$\Delta \text{SNR}_{\text{intellig}} \text{ (dB)} = \sum_{b=1}^B I(\omega_b) \left[ \text{SNR}_{\text{on}}^{\text{on}}(\omega_b) - \text{SNR}_{\text{off}}^{\text{off}}(\omega_b) \right], \quad (31)$$

where  $\text{SNR}_{\text{on}}^{\text{on}}(\omega_b)$  and  $\text{SNR}_{\text{off}}^{\text{off}}(\omega_b)$  denote the SNRs in the  $b$ -th band of the inner error microphone signal (with control) and the leakage (without control), respectively.

In addition, we considered the narrowband perceptual evaluation of speech quality (PESQ) [33] and the extended short-term objective intelligibility (ESTOI) [34] metrics using  $x_{\text{ref},s}(n - \Delta)$  as the reference signal. We evaluated the PESQ and ESTOI differences between the leakage (without control) and the inner error microphone signal (with control).

### 6.3. Results and discussion

Figure 3 compares the performance of the conventional ANC, the hard-constrained SSANC, and the proposed soft-constrained SSANC for different values of the trade-off parameter  $\mu$  ( $-5 \leq \log_{10} \mu \leq 0.5$ ). For the conventional ANC, both desired speech and noise components are minimized at the inner error microphone, yielding the maximum noise reduction (21.6 dB) but also significant speech distortion (0 dB). This leads to a poor SNR improvement (4.4 dB), a drop in the PESQ score ( $-0.12$ ), and a minor ESTOI improvement (0.06). In contrast, the hard-constrained SSANC system almost perfectly preserves the speech component of the reference microphone signal. This results in minimal speech distortion ( $-31.5$  dB) while reducing noise by 12.1 dB. An SNR improvement of 8.4 dB, a PESQ improvement of 0.12, and an ESTOI improvement of 0.33 are obtained.



**Fig. 3:** Noise reduction, intelligibility-weighted spectral distortion, intelligibility-weighted SNR improvement, narrowband PESQ improvement, and ESTOI improvement for conventional ANC (left column,  $\mu = 0$ ), the hard-constrained SSANC (right column,  $\mu \rightarrow \infty$ ), and the proposed soft-constrained SSANC for different values of the trade-off parameter  $\mu$ .

The proposed soft-constrained SSANC system bridges the gap between both extreme cases. For very small values of  $\mu$ , it behaves like the conventional ANC system, whereas for large  $\mu$  it approaches the hard-constrained SSANC system, thus confirming the frequency-domain relationships in (28). More importantly, there is a wide range of parameter values yielding a larger noise reduction than the hard-constrained case without causing excessive speech distortion. For example, setting  $\log_{10} \mu = -2$  ( $\mu = 0.01$ ) yielded a noise reduction of 20.7 dB and a speech distortion of  $-14.7$  dB, corresponding to an SNR improvement of 17.2 dB, a PESQ improvement of 0.54, and an ESTOI improvement of 0.39, outperforming hard-constrained SSANC by 8.8 dB in terms of SNR improvement, 0.42 in terms of PESQ improvement, and 0.06 in terms of ESTOI improvement. These results demonstrate that the soft-constrained SSANC not only generalizes existing approaches but also provides a practical method for improving speech quality and intelligibility<sup>1</sup>.

### 7. CONCLUSION

In this paper, we proposed a soft-constrained SSANC system employing a tunable parameter to trade off between speech distortion and noise reduction. Through a frequency-domain analysis it was shown that the proposed system is equivalent to conventional ANC and hard-constrained SSANC for extreme values of the trade-off parameter, which was confirmed by simulations in the time domain for a pair of open-fitting hearables. The results demonstrated that there exists a wide range of values for the trade-off parameter leading to better SNR improvement and enhanced speech quality and intelligibility than the hard-constrained configuration. Future work includes evaluating the trade-off parameter in different acoustic scenarios.

<sup>1</sup>Audio examples can be found at <https://uol.de/p113620>.

## 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, pp. 1–21, 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] 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.
- [7] 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.
- [8] 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.
- [9] P. R. Benoit, R. Roden, M. Blau, and S. Doclo, "Optimization of a fixed virtual sensing feedback ANC controller for in-ear headphones with multiple loudspeakers," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022, pp. 8717–8721.
- [10] B. Van Veen and K. Buckley, "Beamforming: a versatile approach to spatial filtering," *IEEE ASSP Magazine*, vol. 5, no. 2, pp. 4–24, 1988.
- [11] 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.
- [12] 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.
- [13] R. Serizel, M. Moonen, J. Wouters, and S. H. Jensen, "Output SNR analysis of integrated active noise control and noise reduction in hearing aids under a single speech source scenario," *Signal Processing*, vol. 91, no. 8, pp. 1719–1729, 2011.
- [14] —, "A speech distortion weighting based approach to integrated active noise control and noise reduction in hearing aids," *Signal Processing*, vol. 93, no. 9, pp. 2440–2452, 2013.
- [15] 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.
- [16] —, "Theoretical performance analysis of ANC-motivated noise reduction algorithms for open-fitting hearing aids," in *Proc. International Workshop on Acoustic Signal Enhancement (IWAENC)*, Aachen, Germany, 2012, pp. 1–4.
- [17] C.-Y. Ho, K.-K. Shyu, C.-Y. Chang, and S. M. Kuo, "Integrated active noise control for open-fit hearing aids with customized filter," *Applied Acoustics*, vol. 137, pp. 1–8, 2018.
- [18] H. Zhang, J. Zhang, F. Ma, P. N. Samarasinghe, and H. Sun, "A time-domain multi-channel directional active noise control system," in *Proc. European Signal Processing Conference (EUSIPCO)*, Helsinki, Finland, 2023, pp. 376–380.
- [19] H. Zhang, H. J. Sun, J. A. Zhang, P. Samarasinghe, and Y. A. Zhang, "A spherical-harmonic domain selective spatial active noise control system based on sound field reproduction," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Hyderabad, India, 2025, pp. 1–5.
- [20] T. Xiao and S. Doclo, "Effect of target signals and delays on spatially selective active noise control for open-fitting hearables," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Seoul, Republic of Korea, 2024, pp. 1056–1060.
- [21] —, "Spatially selective active noise control for open-fitting hearables with acausal optimization," in *Proc. Forum Acusticum Euronoise 2025*, Málaga, Spain, Jun. 2025.
- [22] S. Doclo and M. Moonen, "GSVD-based optimal filtering for single and multimicrophone speech enhancement," *IEEE Transactions on Signal Processing*, vol. 50, no. 9, pp. 2230–2244, 2002.
- [23] A. Spriet, M. Moonen, and J. Wouters, "Spatially pre-processed speech distortion weighted multi-channel Wiener filtering for noise reduction," *Signal Processing*, vol. 84, no. 12, pp. 2367–2387, 2004.
- [24] S. Nordholm, H. Q. Dam, N. Grbic, and S. Y. Low, "Adaptive microphone array employing spatial quadratic soft constraints and spectral shaping," *Signals and Communication Technology*, pp. 229–246, 2005.
- [25] S. Doclo, A. Spriet, J. Wouters, and M. Moonen, "Frequency-domain criterion for the speech distortion weighted multichannel Wiener filter for robust noise reduction," *Speech Communication*, vol. 49, no. 7, pp. 636–656, 2007.
- [26] S. Braun, K. Kowalczyk, and E. A. P. Habets, "Residual noise control using a parametric multichannel Wiener filter," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, South Brisbane, QLD, Australia, 2015, pp. 360–364.
- [27] H. Schepker, S. Nordholm, and S. Doclo, "Acoustic feedback suppression for multi-microphone hearing devices using a soft-constrained null-steering beamformer," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 929–940, 2020.
- [28] 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, pp. 1–9.
- [29] 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, no. 2, pp. 1–16, 2021.
- [30] C. Veaux, J. Yamagishi, and K. MacDonald, "CSTR VCTK corpus: English multi-speaker corpus for CSTR voice cloning toolkit," 2017. [Online]. Available: <https://doi.org/10.7488/ds/2645>
- [31] 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.
- [32] Acoustical Society of America (ASA), "Methods for Calculation of the Speech Intelligibility Index," American National Standards Institute (ANSI), ANSI/ASA S3.5-1997 Standard, 1997.
- [33] 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.
- [34] J. Jensen and C. H. Taal, "An algorithm for predicting the intelligibility of speech masked by modulated noise maskers," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 11, pp. 2009–2022, 2016.