Empirical Analysis of Waterbed-like Performance Trade-Off for a Feedforward Occlusion Effect Reduction System

Christoph Weyer, Florian Hilgemann, and Peter Jax
Institute of Communication Systems (IKS), RWTH Aachen University, Germany
{weyer, hilgemann, jax}@iks.rwth-aachen.de

***Abstract*—While wearing earbuds, the wearers often perceive their own voice as boomy. This so-called occlusion effect (OE) poses a major problem. Various countermeasures based on active noise cancellation (ANC) have been proposed. In a previous publication, we investigated the use of a vibration sensor in a time-invariant feedforward ANC scheme to reduce the OE. However, a potential drawback is the leakage of sensor noise into the ear canal. In this publication, we employ an improved filter design method which allows us to constrain the noise leakage. We show empirically that constraining the noise leakage also results in limited OE reduction performance. This poses a trade-off similar to what is known for feedback systems as the waterbed effect. Furthermore, we show that the trade-off is rooted in the fact that we can only use causal filters in practice. Lastly, we give an example on how the improved filter design method, further extended by an outer microphone as reference sensor, can be of use in an application-oriented filter design. The analysis is done via simulations based on measurements with a hardware prototype.**

***Index Terms*—Hearables, occlusion effect reduction, active noise cancellation, feedforward filter design, waterbed effect**

I. Introduction

Earbuds are becoming ever more ubiquitous, however, their wearers often perceive their own voice as well as other body-conducted sounds, such as footsteps, as boomy. This phenomenon, called the occlusion effect (OE), is caused by two mechanisms: Firstly, air-conducted speech components are attenuated by the earbud, mostly affecting higher frequencies. Secondly, the predominantly low frequency body-conducted speech components are amplified as the vibrating ear canal walls can better excite the air volume in the occluded ear canal [1, p. 26]. Active noise cancellation (ANC) techniques can be used to reduce the OE, which is promising, as many modern earbuds are capable of ANC. Different approaches, mostly focusing on counteracting the amplification of low frequency components, have been described in the literature. They mainly employ feedback (FB) control using a microphone facing into the ear canal, e.g. [1]–[5]. However, feedforward (FF) control using an outward-facing microphone has also been proposed [6], [7]. We recently investigated a novel FF system using a vibration sensor [8], an approach that has so far only been proposed in patents [9], [10]. The vibration sensor is capable of directly sensing body-conducted speech components. Thus, it can be used for generating a cancellation signal to counteract the amplification of low frequency components in the ear canal. Additionally, vibration sensors are typically more robust against wind noise, but also air-conducted sound, which offers clear advantages compared to microphone-based systems.

Among other limitations, e.g. due to the danger of instability, the performance of FB-based control systems is limited by the waterbed effect [11, p. 164]. This effect causes a desired amount of disturbance rejection in one frequency region to be accompanied by an equal amount of amplification in another region, thus, posing a fundamental design trade-off. However, similar effects have also been described for FF systems, i.e. in the absence of FB loops, e.g. [12], [13].

We already noted in [8] that the proposed system tends to leak undesired signal components into the ear canal above 1 kHz. There, we regularized the FF filter for higher frequencies to limit this leakage. In this publication, we show that there is a fundamental design trade-off between noise leakage and OE reduction performance inherent to the system, which is similar to the waterbed effect. We do so using simulations based on acoustic measurements with a hardware prototype of an in-ear ANC headphone. Afterwards, we conduct an analysis of the trade-off, utilizing an improved filter design which enables us to constrain the noise leakage in a targeted manner without the need for manually adjusting regularization filters. Furthermore, the outer microphone is used as an additional reference sensor, mainly allowing us to also counteract the attenuation of high frequency components. We use a signal-based filter design as the underlying system possibly depends on the physiology of the wearer as well as wearing conditions, and its description is thus non-trivial. While the sensor in [8] was attached to the skin of the user, we extend our research to an earbud-mounted vibration sensor, as it suits itself better for practical application. This poses potentially more problems with feedback from the speaker to the vibration sensor, which however are out of scope for this paper.

II. Theory

A. System Model

Fig. 1 shows the system model as well as a sketch of the underlying physical components. The earbud contains a loudspeaker, a vibration sensor, a microphone facing into the ear canal, subsequently called *inner* microphone, as well as a microphone facing outwards, subsequently called *outer* microphone. All components are connected to a digital signal

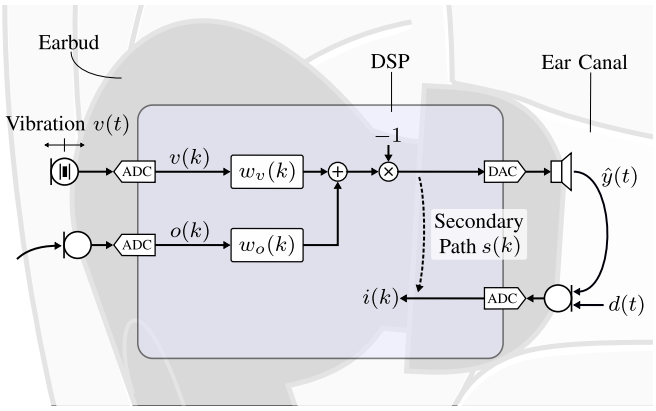

Fig. 1: System model with underlying physical components.

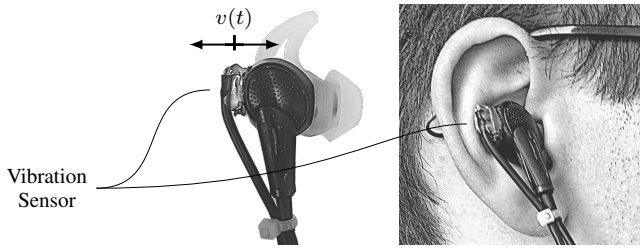

Fig. 2: Bose QC20 earbud with attached vibration sensor.

processing (DSP) platform. In the following, let k be the discrete time index, while t is the continuous time. The vibration sensor signal $v(k)$ captures the digitized earbud vibration in one axis. The orientation of this axis relative to the earbud is indicated in Fig. 2. The outer microphone signal $o(k)$ captures the digitized sound pressure outside the ear canal. Both signals, $v(k)$ and $o(k)$, are filtered by a digital FF filter, which we describe by their impulse responses (IRs) $w_v(k)$ and $w_o(k)$, respectively. Furthermore, the inner microphone signal $i(k)$ captures the digitized sound pressure inside the ear canal, given as the sum

$$i(k) = \hat{y}(k) + d(k), \tag{1}$$

where $\hat{y}(k)$ corresponds to the signal component caused by the speaker, and $d(k)$ corresponds to all remaining components. I.e., $d(k)$ is the signal *before* and $i(k)$ the signal *after* acoustic addition of the speaker signal. In turn, the speaker signal at the inner microphone is given by

$$\hat{y}(k) = -s(k) * \left[w_v(k) * v(k) + w_o(k) * o(k)\right], \tag{2}$$

where $s(k)$ is the IR of the secondary path, which describes the transmission from the digital input of the speaker to the digital output of the inner microphone, as indicated in Fig. 1.

B. Metrics

We quantify how the system changes signal power in the ear canal using the power gain (PG)

$$G(\Omega) = \sqrt{\Phi_i(\Omega) \,/\, \Phi_d(\Omega)}, \tag{3}$$

where $\Phi_i(\Omega)$ and $\Phi_d(\Omega)$ are the long-term average power spectra (PS) of the inner microphone signal $d(k)$ before and $i(k)$ after addition of the speaker signal, respectively, and $\Omega = 2\pi f / f_\mathrm{s}$ is the angular frequency relative to sampling rate f_s, depending on frequency f. Furthermore, we quantify the OE *before* addition of the speaker signal as $\mathrm{OE}_d(\Omega) = \sqrt{\Phi_d(\Omega)/\Phi_{o_\mathrm{c}}(\Omega)}$, as well as the OE *after* addition of the speaker signal as $\mathrm{OE}_i(\Omega) = \sqrt{\Phi_i(\Omega)/\Phi_{o_\mathrm{c}}(\Omega)}$. Here, $\Phi_{o_\mathrm{c}}(\Omega)$ is the long-term average PS of the calibrated outer microphone signal $o_\mathrm{c}(k)$. We obtain the latter as $o_\mathrm{c}(k) = \gamma\, o(k)$, where γ is the gain corresponding to the dB-level average of $\Phi_d(\Omega)/\Phi_o(\Omega)$ for a diffuse field recording, in the frequency range from $50\,\mathrm{Hz}$ to $1\,\mathrm{kHz}$. Now, if the proposed system reverses the change in signal power in the ear canal caused by the occlusion, i.e. $G(\Omega) \approx 1/\mathrm{OE}_d(\Omega)$, this indicates good performance. Thus, we will use $1/\mathrm{OE}_d(\Omega)$ as a target for the PG later on when we evaluate performance.

C. Filter Design

Now, we want to find filter IRs $w_v(k)$ and $w_o(k)$ so the OE is reduced, i.e. $i(k) \approx i_\mathrm{open}(k)$, where $i_\mathrm{open}(k)$ is the signal as it would be recorded in the unoccluded ear canal. Using (1), we derive the desired speaker signal as

$$i(k) \approx i_\mathrm{open}(k) \;\Rightarrow\; \hat{y}(k) \approx i_\mathrm{open}(k) - d(k) = y(k), \tag{4}$$

where we define $y(k)$ as the target for the speaker signal $\hat{y}(k)$ to achieve the outlined goal. In the following, we choose different design goals depending on the experiment. Thus, we state the solution depending on the generalized target $y(k)$. Writing the filter IRs as column vectors $\mathbf{w}_v, \mathbf{w}_o \in \mathbb{R}^L$ of length L, we find the filters that minimize the squared error between the speaker signal $\hat{y}(k)$ and the target signal $y(k)$, including regularization, by minimizing the cost

$$c(\mathbf{w}) = -2\,\mathbf{w}^\mathrm{T}\boldsymbol{\varphi} + \mathbf{w}^\mathrm{T}\,\underline{\boldsymbol{\varphi}}\,\mathbf{w} \tag{5}$$

with respect to the stacked filter vector $\mathbf{w} = \begin{bmatrix}\mathbf{w}_v^\mathrm{T} & \mathbf{w}_o^\mathrm{T}\end{bmatrix}^\mathrm{T} \in \mathbb{R}^{2L}$. Here, $\boldsymbol{\varphi} \in \mathbb{R}^{2L}$ is the stacked cross-correlation vector and $\underline{\boldsymbol{\varphi}} \in \mathbb{R}^{2L\times 2L}$ is the stacked correlation matrix, given as

$$\boldsymbol{\varphi} = \begin{bmatrix}\boldsymbol{\varphi}_{\tilde{v}y} \\ \boldsymbol{\varphi}_{\tilde{o}y}\end{bmatrix} \;\text{ and }\; \underline{\boldsymbol{\varphi}} = \begin{bmatrix}\underline{\boldsymbol{\varphi}}_{\tilde{v}\tilde{v}} & \underline{\boldsymbol{\varphi}}_{\tilde{v}\tilde{o}} \\ \underline{\boldsymbol{\varphi}}_{\tilde{v}\tilde{o}}^\mathrm{T} & \underline{\boldsymbol{\varphi}}_{\tilde{o}\tilde{o}}\end{bmatrix} + \begin{bmatrix}\underline{\boldsymbol{\varphi}}_{r_v r_v} & 0 \\ 0 & \underline{\boldsymbol{\varphi}}_{r_o r_o}\end{bmatrix}. \tag{6}$$

In turn, $\boldsymbol{\varphi}_{\tilde{v}y}, \boldsymbol{\varphi}_{\tilde{o}y} \in \mathbb{R}^L$ are the cross-correlation vectors of the signals $\tilde{v}(k)$ and $\tilde{o}(k)$ with the target $y(k)$, respectively, and $\underline{\boldsymbol{\varphi}}_{\tilde{v}\tilde{v}}, \underline{\boldsymbol{\varphi}}_{\tilde{o}\tilde{o}} \in \mathbb{R}^{L\times L}$ are the auto-correlation matrices and $\underline{\boldsymbol{\varphi}}_{\tilde{v}\tilde{o}} \in \mathbb{R}^{L\times L}$ is the cross-correlation matrix between $\tilde{v}(k)$ and $\tilde{o}(k)$. Here, $\tilde{v}(k)$ and $\tilde{o}(k)$ are the reference signals which are pre-filtered with the sign-inverted secondary path IR in a filtered-X scheme, i.e. $\tilde{v}(k) = -s(k) * v(k)$ and $\tilde{o}(k) = -s(k) * o(k)$. Lastly, $\underline{\boldsymbol{\varphi}}_{r_v r_v}, \underline{\boldsymbol{\varphi}}_{r_o r_o} \in \mathbb{R}^{L\times L}$ are the auto-correlation matrices of the regularization filters $r_v(k)$ and $r_o(k)$, respectively. Note that (5) and (6) constitute a standard multichannel inverse filtering problem similar to [14] but using a signal-based formulation.

To tackle the noise leakage, we constrain the filters during design, a concept which has in general been described for

control applications in e.g. [15]. As the leakage of incoherent noise through the FF filters leads to an increased PG, we attempt to constrain the PG, as defined in (3), during optimization. As noise leakage is, in practice, mainly a problem for the vibration sensor signal, we assume $w_o(k) = 0$ for the computation of the constraint, thus focusing on the output $v(k) * w_v(k)$ of the vibration sensor FF filter. During optimization, we approximate the PG as ratio of the discrete-time Fourier transforms (DTFTs) of the inner microphone signals after and before addition of the speaker output, yielding the approximated power gain

$$\mathcal{G}_v(\Omega) = |D(\Omega) + \tilde{V}(\Omega)\,\mathbf{q}(\Omega)^{\mathrm{T}}\mathbf{w}_v| \;/\; |D(\Omega)|, \tag{7}$$

where $\mathbf{q}(\Omega) = \begin{bmatrix}1 & e^{-\mathrm{j}\Omega} & e^{-\mathrm{j}2\Omega} & \dots & e^{-\mathrm{j}(L-1)\Omega}\end{bmatrix}^{\mathrm{T}} \in \mathbb{C}^L$ is the fourier transform vector, with the product $\mathbf{q}(\Omega)^{\mathrm{T}}\mathbf{w}_v$ yielding the DTFT of $\mathbf{w}_v$ at Ω, and $D(\Omega)$ and $\tilde{V}(\Omega)$ are DTFTs of the inner microphone signal before cancellation $d(k)$ and the vibration sensor signal pre-filtered with the secondary path $\tilde{v}(k)$, respectively. Note that the absolute value of $I(\Omega)$ and $\tilde{V}(\Omega)$ are smoothed in $1/3$ octave bands. During optimization, we constrain $\mathcal{G}_v(\Omega)$ using the user-selectable upper bound $\overline{\mathcal{G}}_v(\Omega)$ for a set $\mathcal{O}$ of normalized angular frequencies, i.e.

$$\mathcal{G}_v(\Omega) \leq \overline{\mathcal{G}}_v(\Omega) \text{ for all } \Omega \in \mathcal{O}. \tag{8}$$

This allows us to limit the noise leakage in a frequency-dependent way, without the need for manually adjusting the regularization filter $r_v(k)$.

In total, this leads to the convex optimization problem

$$\begin{aligned} &\min_{\mathbf{w}} \; -2\,\mathbf{w}^{\mathrm{T}}\boldsymbol{\varphi} + \mathbf{w}^{\mathrm{T}}\,\underline{\boldsymbol{\varphi}}\,\mathbf{w} \\ &\text{s.t.} \quad \mathcal{G}_v(\Omega) \leq \overline{\mathcal{G}}_v(\Omega) \text{ for all } \Omega \in \mathcal{O}. \end{aligned} \tag{9}$$

III. Experimental Setup

We used a pair of modified Bose QC20 earbuds to obtain data for our experiments. Fig. 2 shows the right modified earbud. The internal ANC processing was disabled, and we directly accessed the microphones and speakers. A Knowles BU-23842 accelerometer was attached to each earbud using hot glue. All sensors and speakers were connected to a DSP platform based on an Analog Devices ADAU 1787 as well as an ADAU 1777 audio codec, which were in turn connected to a computer via sound card. We repeated the following measurements for six subjects, five male and one female, and for two refits of the earbuds per subject. We recorded all sensor signals for the left and right earbud while the subjects read a segment of *the rainbow passage* as described in [16, p. 124-139]. The speech recordings had an average length of 35 s. Furthermore, we identified the secondary path IRs $s(k)$ for each fit using a 10 s long exponential sweep. All recordings were done at sampling rate $f_\mathrm{s} = 48\,\mathrm{kHz}$. Combining data from all recordings and left and right side, this yields a set of 24 samples in total.

If not stated otherwise, we used the following practices throughout the paper and the experiments: We always considered all 24 data samples. Furthermore, the default length for the designed FF filters $w_v(k)$ and $w_o(k)$ was $L = 1024$ taps, which corresponds to a time interval of around 21 ms. We solved (9) using a set $\mathcal{O}$ of angular frequency points, corresponding to uniformly spaced frequencies from 20 Hz on with a spacing of $f_\mathrm{s}/L/5$, corresponding to roughly 9.4 Hz for the default filter length, up to and including $f_\mathrm{s}/2$. The dB-levels of all shown spectral measures were smoothed in $1/3$ octave bands for the sake of clear presentation. If we show averaged data, we indicate the data on which the averages are based on with thinner lines of similar color for completeness.

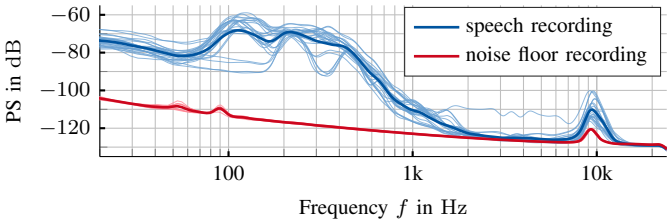

Fig. 3: Signal characteristics of vibration sensor setup.

IV. Results

A. Signal-to-Noise Ratio of Vibration Sensor Setup

First, we illustrate the signal-to-noise ratio (SNR) of our vibration sensor setup for the case of speech excitation. To characterize the noise, we recorded the vibration sensor signals four times for 40 s in an acoustically insulated room, while the earbuds where not worn. Fig. 3 shows the dB-level average of the PS over both sides and over all noise (——) recordings and over all speech (——) recordings from Sec. III. We observe that the SNR of the vibration sensor setup is around zero above 2 kHz, and thus, the vibration sensor signal is unlikely to be useful for counteracting the OE in this frequency region. Note, that the peak around $f = 9\,\mathrm{kHz}$ reflects the sensor resonance.

B. Description of Trade-Off

Now, we show that a trade-off between cancellation performance and noise leakage is inherent to the system. We designed FF filters $w_v(k)$ and simulated the resulting inner microphone signal $i(k)$ using the data described in Sec. III. To assess the trade-off independent of system variations, e.g. due to inter-person variability, we designed 24 separate filters, i.e. one filter for each pair of recording and secondary path, and simulated $i(k)$ for each filter using the recording and path it was designed on. We chose $i(k) \approx 0$ as our design goal, i.e. maximum possible attenuation, to rule out that undesired amplification is caused by the system trying to boost high frequency components, which would be necessary to fully counteract the OE. The chosen goal results in the target signal being $y(k) = -d(k)$, i.e. $d(k)$ acts as a disturbance to be cancelled. Moreover, as we focus on the properties of the vibration sensor signal, we set the outer microphone signal to $o(k) = 0$ in the design and disregarded the filter output $w_o(k) * o(k)$. We did not employ regularization as this acts as an additional, soft constraint in the filter design, and we want to focus on the effect of (8) here.

To outline the noise leakage problem, we first designed filters without constraint. Fig. 4 shows the resulting dB-level averages

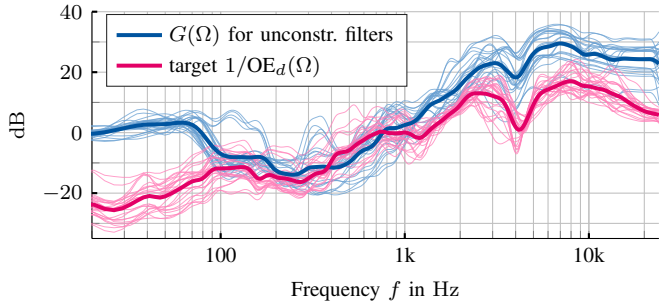

Fig. 4: Performance of unconstrained filters.

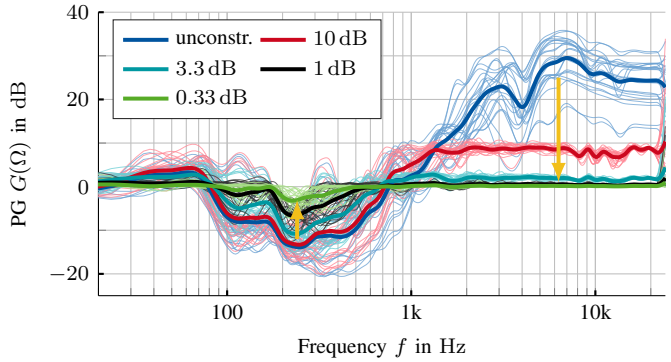

Fig. 5: Influence of constraint $\overline{\mathcal{G}}_v(\Omega)$ on performance.

of the PGs over all recordings (——). Furthermore, we indicate the dB-level average of $1/\mathrm{OE}_d(\Omega)$ over all recordings (——), which acts as a target for the PG as outlined in Sec. II-B. The averages indicate that we can counteract a significant portion of the OE-induced overshoot above 80 Hz. However, we also observe a considerable amplification below 80 Hz and a rather strong amplification of up to around 30 dB above 800 Hz. Given the low SNR above 2 kHz, as outlined in Sec. IV-A, this results in a lot of noise being leaked into the ear canal. This is also clearly audible in the inner microphone signals $i(k)$ that result from the simulation.

Next, we designed filters with different constraints $\overline{\mathcal{G}}_v(\Omega)$ to limit the noise leakage. We chose $\overline{\mathcal{G}}_v(\Omega)$ ranging from 0.33 dB, i.e. the constraint is very restrictive, to 10 dB, i.e. the constraint is not very restrictive. Fig. 5 shows the dB-level averages of the resulting PGs over all recordings for

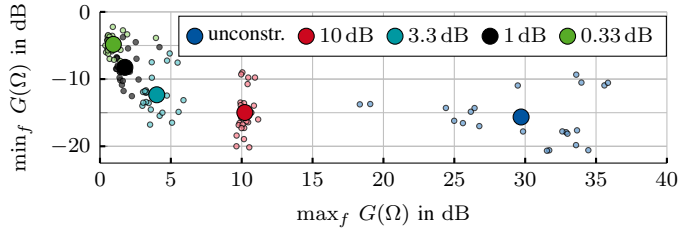

Fig. 6: Maximum attenuation, i.e. $\min_f G(\Omega)$, over maximum noise leakage, i.e. $\max_f G(\Omega)$, for different constraints $\overline{\mathcal{G}}_v(\Omega)$.

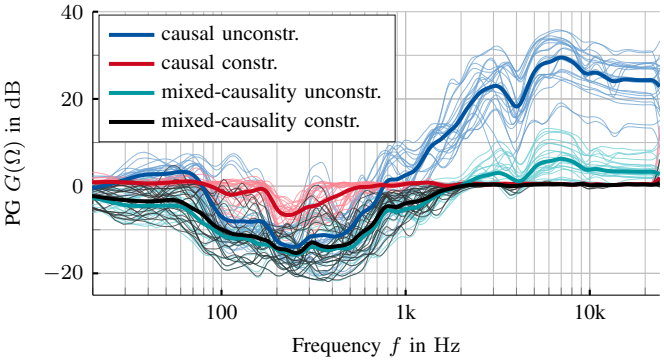

Fig. 7: Influence of causality on performance.

the different constraints. The unconstrained case is indicated again for reference. We observe that decreasing the upper bound $\overline{\mathcal{G}}_v(\Omega)$ decreases the noise leakage above 800 Hz as intended. Note that we only constrained the PG as defined in (3) indirectly, i.e. using (7). Thus, the observed PGs can slightly exceed the constraint. However, there is still a clear correspondence between decreasing the constraint $\overline{\mathcal{G}}_v(\Omega)$ and a decrease in the observed PG visible. We see that a decrease in noise leakage above 800 Hz also results in a decrease in attenuation below 800 Hz, e.g. for $\overline{\mathcal{G}}_v(\Omega) = 1\,\mathrm{dB}$ (——), we observe almost no visible noise leakage anymore, but the maximum attenuation also drops. We conclude that there is a trade-off between cancellation performance and noise leakage. This relationship is indicated in Fig. 5 by the yellow arrows. This trade-off is similar to the so-called waterbed effect present in feedback systems [11, p. 164]

We further visualize the relation between maximum attenuation and maximum noise leakage in Fig. 6. To do so, we computed the maximum attenuation, i.e. $\min_f G(\Omega)$, and the maximum noise leakage, i.e. $\max_f G(\Omega)$, for the $1/3$ octave band smoothed PGs from this experiment. For the computation of the minima and maxima, we did not consider frequencies below 20 Hz as the PG was not constrained here during design, but also not above 20 kHz as the PG tends to contain artifacts from the smoothing there. We indicate the centroid of all points corresponding to a particular upper bound $\overline{\mathcal{G}}_v(\Omega)$ with a thicker point. The centroids clearly confirm the described trade-off.

C. Influence of Causality Constraint on Trade-Off

While only causal filters can be implemented in practice, we now relax the causality constraint to evaluate its influence on the trade-off. To do so, we repeated the simulations from Sec. IV-B, however, we also considered mixed-causality filters whose IRs $w_v(k)$ can have non-zero coefficients for $-1024 \leq k \leq 1023$, in contrast to $0 \leq k \leq 1023$ for causal filters. Also in contrast to Sec. IV-B, we focused only on filters designed without a constraint and with the constraint $\overline{\mathcal{G}}_v(\Omega) = 1\,\mathrm{dB}$. In total, the combination of causality and PG constraints yields four combinations of design constraints.

Fig. 7 shows the dB-level averages of the resulting PGs over all recordings for the four combinations. For the causal filters

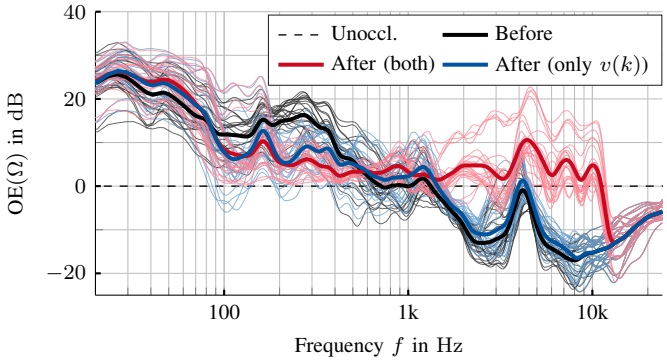

Fig. 8: Effect of system on OE.

(——, ——), we observe the same trade-off as earlier. However, for the mixed-causality filters (——,——), we observe less noise leakage for the unconstrained filters (——), as well as similar performance for constrained and unconstrained filters. The latter means that the severity of the trade-off is reduced significantly. Thus, the trade-off seems to be rooted in the fact that applicable filters need to be causal.

D. Practical Example

Lastly, we want to give a more application-oriented example of how the design method described in Sec. II-C could be applied. To fully counteract the OE, we now chose $i(k) \approx o_\mathrm{c}(k-24)$ as design goal, resulting in $y(k) = o_\mathrm{c}(k-24) - d(k)$. We delay $o_\mathrm{c}(k)$ for the target to avoid problems because of the non-minimum phase nature of the secondary path, which implicitly needs to be inverted in the design [17]. Thus, we delayed $o_\mathrm{c}(k)$ by around $0.5\,\mathrm{ms}$ in accordance with [18]. We now also included the outer microphone signal $o(k)$ as additional reference signal. We chose the constraint $\overline{\mathcal{G}}_v(\Omega) = 3.3\,\mathrm{dB}$. We chose the regularization filters $r_v(k)$ and $r_o(k)$ in a way that low gain is incentivized for $w_v(k)$ above $8\,\mathrm{kHz}$ and for $w_o(k)$ above $12\,\mathrm{kHz}$.

Fig. 8 shows the dB-level average of the OE over all recordings before (——) and after (——) addition of the speaker signal. To disentangle the contributions of both reference sensors, we also simulated $i(k)$ by only applying the vibration sensor's FF filters $w_v(k)$ output. Note, that no separate design of the filters $w_v(k)$ was done for this case. (——) shows the resulting OE. We observe that the OE is on average reduced by the proposed system, indicated by (——) being closer to $0\,\mathrm{dB}$, i.e. the unoccluded case. Furthermore, the constraint acts as intended and limits the leakage of noise which is indicated by (——) only slightly exceeding (——) for frequencies below $80\,\mathrm{Hz}$ and above $800\,\mathrm{Hz}$. Note here that the constraint enables us to relatively evenly distribute the noise leakage among the spectrum, which should be perceptually advantageous, without the need for manually adjusting regularization filters. Moreover, the outer microphone not only enables counteracting the attenuation of high frequency components, it also contributes somewhat to reducing the overshoot of low frequency components.

V. Conclusion

In this publication, we showed that a trade-off between achievable cancellation performance and noise leakage exists in a proposed vibration sensor-based FF OE reduction system. We analyzed the trade-off, and showed that it implies restrictions in the filter design similar to the waterbed effect known mainly from FB control. Further analysis showed that the trade-off is rooted in the causality constraint of the system. We employed a constraint in the filter design to analyze the problem, but also gave an example of how the resulting filter design method could be beneficial in practice.

References

[1] S. Liebich, "Active noise and occlusion effect cancellation in headphones and hearing aids," Ph.D. dissertation, RWTH Aachen University, Aachen, Germany, 2020.

[2] J. P. Mejia, H. A. Dillon, and M. J. A. Fisher, "Active cancellation of occlusion: An electronic vent for hearing aids and hearing protectors," *The Journal of the Acoustical Society of America*, vol. 124, no. 1, pp. 235–240, Jul. 2008.

[3] M. Sunohara, K. Watanuki, and M. Tateno, "Occlusion reduction system for hearing aids using active noise control technique," *Acoustical Science and Technology*, vol. 35, no. 6, pp. 318–320, Jun. 2014.

[4] R. C. Borges, M. H. Costa, J. A. Cordioli, and L. F. C. Assuiti, "An adaptive occlusion canceller for hearing aids," in *21st European Signal Processing Conference (EUSIPCO 2013)*, Sep. 2013.

[5] T. S. Zurbrügg, "Active control mitigating the ear canal occlusion effect caused by hearing aids," Ph.D. dissertation, EPFL, Lausanne, Switzerland, 2014.

[6] R. C. Borges and M. H. Costa, "A feed forward adaptive canceller to reduce the occlusion effect in hearing aids," *Computers in Biology and Medicine*, vol. 79, pp. 266–275, Dec. 2016.

[7] B. C. Bispo and R. C. Borges, "A cepstral method to estimate the stable optimal solution for feedforward occlusion cancellation in hearing aids," *Journal of Communication and Information Systems*, vol. 35, no. 1, pp. 113–123, May 2020.

[8] C. Weyer and P. Jax, "Occlusion effect reduction using a vibration sensor," in *2022 International Workshop on Acoustic Signal Enhancement (IWAENC)*, Sep. 2022.

[9] P. Isberg, A. Petef, O. Moliner, and O. Thörn, "Controlling own-voice experience of talker with occluded ear," US Patent 9 949 048 B2, Apr., 2018.

[10] T. P. Hua, E. B. Andersen, T.-D. W. Saux, and S. C. Grinker, "Headphone transparency, occlusion effect mitigation and wind noise detection," US Patent 10 657 950 B2, May, 2020.

[11] S. Skogestad and I. Postlethwaite, *Multivariable Feedback Control - Analysis and Design*, 2nd ed. John Wiley & Sons, 2001.

[12] M. F. Heertjes, B. Temizer, and M. Schneiders, "Self-tuning in master–slave synchronization of high-precision stage systems," *Control Engineering Practice*, vol. 21, no. 12, pp. 1706–1715, Dec. 2013.

[13] M. A. Beijen, M. F. Heertjes, H. Butler, and M. Steinbuch, "Performance trade-offs in disturbance feedforward compensation of active hard-mounted vibration isolators," in *2015 American Control Conference (ACC)*, Jul. 2015, pp. 2149–2154.

[14] O. Kirkeby and P. A. Nelson, "Digital filter design for inversion problems in sound reproduction," *Journal of the Audio Engineering Society*, vol. 47, no. 7/8, pp. 583–595, Jul. 1999.

[15] S. Boyd, C. Baratt, and S. Norman, "Linear controller design: Limits of performance via convex optimization," *Proceedings of the IEEE*, vol. 78, no. 3, pp. 529–574, Mar. 1990.

[16] G. Fairbanks, *Voice and Articulation Drillbook*, 2nd ed. New York: Harper & Row, 1960.

[17] S. Liebich, R. Brandis, J. Fabry, P. Jax, and P. Vary, "Active occlusion cancellation with hear-through equalization for headphones," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Apr. 2018, pp. 241–245.

[18] F. Denk, H. Schepker, S. Doclo, and B. Kollmeier, "Equalization filter design for achieving acoustic transparency in a semi-open fit hearing device," in *Speech Communication; 13th ITG-Symposium*, Oct. 2018.