Adaptive Sparse Linear Prediction in Fixed-Filter ANC Headphone Applications for Multi-Speaker Speech Reduction

Iotov, Yurii; Nørholm, Sidsel Marie; Belyi, Valiantsin; Christensen, Mads Græsbøll

Published in:
Proceedings of the 2023 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, WASPAA 2023

DOI (link to publication from Publisher):
10.1109/WASPAA58266.2023.10248065

Publication date:
2023

Document Version
Accepted author manuscript, peer reviewed version

ADAPTIVE SPARSE LINEAR PREDICTION IN FIXED-FILTER ANC HEADPHONE APPLICATIONS FOR MULTI-SPEAKER SPEECH REDUCTION

Yurii Iotov[*1,2], *Sidsel Marie Nørholm*[2], *Valiantsin Belyi*[2], *Mads Græsbøll Christensen*[1]

[1] Audio Analysis Lab, CREATE, Aalborg University, Denmark, {yio, mgc}@create.aau.dk
[2] GN Audio A/S, Ballerup, Denmark, {yiotov, snoerholm, vbelyi}@jabra.com

ABSTRACT

In some cases, speech can be a disturbing source of ambient noise. Active noise control (ANC) systems have difficulties in dealing with speech due to its non-stationary nature and constraints in the ANC system, which require the optimal filters to be non-causal. The non-causality is due to the delay incurred by, e.g., digital processing or acoustic propagation paths. We propose a fixed-filter feedforward ANC system, HOSpLP-ANC, which aims at attenuating voiced speech in, e.g., office environments. It comprises an adaptive high-order sparse linear predictor (HOSpLP) based on the improved proportionate normalised least mean square algorithm to predict speech ahead in time, thus overcoming such delay. Notably, HOSpLP provides high prediction performance of voiced speech by modelling the joint short- and long-term linear prediction scheme, but without using pitch estimation. This can be of particular significance in the case of the complicated multi-pitch estimation scenario. The results show that HOSpLP-ANC outperforms conventional adaptive feedforward ANC for delays in the order of milliseconds in both single- and multi-speaker environments.

Index Terms— Speech attenuation, ANC causality, IPNLMS.

1. INTRODUCTION

Nowadays, the use of ANC technology spans a wide variety of applications [1–4]. Specifically, ANC is becoming more prominent and widespread in consumer electronics, e.g., ANC headphones, headsets and small wireless earbuds. Among the various ambient noise sources we are dealing with in everyday life, speech can be a very disturbing source of ambient noise. For example, in crowded public spaces or open offices, speech may be even more annoying than other types of noise, reducing concentration and productivity. Hence, it is becoming increasingly important that ANC headphones also attenuate human voices effectively. However, speech attenuation by ANC headphones can be quite limited due to the complex nature of speech [5] and constraints in the ANC system.

ANC is based on the principle of acoustic superposition, such that an anti-noise signal with the same amplitude but the opposite phase is generated by a secondary source (e.g., headphone loudspeaker) to cancel unwanted noise at the desired cancellation point, e.g., at the eardrum. To generate an anti-noise signal in adaptive ANC systems, adaptive algorithms such as Filtered-X least mean square (FXLMS) or Filtered-X Normalized LMS (FXNLMS) are commonly used [3, 6–8]. Modern ANC headphones are typically based on either fixed feedforward (FF) or feedback (FB) ANC filter or a combination of both (hybrid ANC) [9].

*The work is supported by the Innovation Fund Denmark, grant no. 9065-00218.

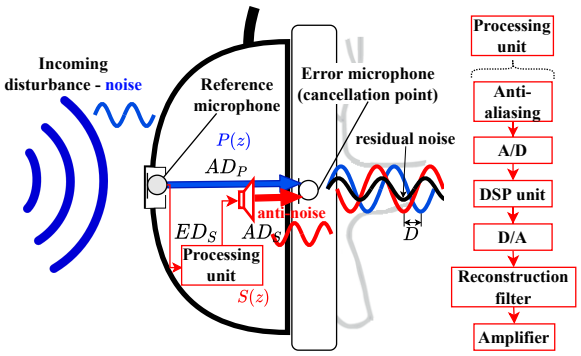

Figure 1: Simplified modelling block diagram of FF ANC headphones with the additional delay $D = ED_S + AD_S - AD_P$ in $S(z)$.

Many factors affect ANC performance in headphones [2,6–16], one of those factors is the causality constraint. For FF ANC headphones, shown in Fig. 1, when the signal propagation delay AD_P between the reference and the error microphone of the primary path $P(z)$ is less than the electric delay ED_S and acoustic propagation delay AD_S in the secondary path $S(z)$, i.e., $AD_P < ED_S + AD_S$, the causality constraint is violated, with the additional delay in $S(z)$ compared to $P(z)$, $D = ED_S + AD_S - AD_P$. The causality constraint might be violated due to the small size of the headphones, making AD_P smaller compared to the combination of AD_S and ED_S. The amount of ED_S depends on the ANC processing unit and its algorithmic design [14–16]. The delay D might also be affected by the direction of the incoming noise [11] and improper headphone fit on the ear. When the causality constraint is violated, it creates the need for prediction to compensate for D [10]. In adaptive ANC, the adaptive algorithm acts as a predictor to find a causal filter [13]. For a fixed-filter ANC system, the occurrence of D cannot be compensated for in the fixed-filter design stage, and, thereby, the performance of such a system will be significantly reduced [10–16]. Hence, it is critical to compensate for the delay D.

Speech tends to be highly non-stationary, and it has a complex structure [5]. Speech sounds can be broadly divided into voiced and unvoiced speech. Voiced speech is the main constituent of speech and normally has higher power than unvoiced speech. Unvoiced speech has a stochastic nature with low correlations and is therefore almost unpredictable [5, 17]. A common approach for speech prediction is linear prediction (LP), the fundamental idea of which is that a speech sample can be approximated as a linear combination of past samples [5]. However, the major contribution to voiced speech prediction performance is mainly due to the most recent 10-12 samples (at a sampling frequency, f_s=8 kHz), and the

samples at the pitch period of speech, T [5]. This corresponds to the short- and long-term correlations of voiced speech, i.e., short- and long-term LP (STP and LTP), the joint modelling of which, namely SLTPj, was proposed in [18]. The SLTPj scheme can also be seen as a high-order sparse filter with the distinguished nonzero regions of taps corresponding to STP and LTP. This inspired the idea of modelling SLTPj using high-order sparse linear prediction (HOSpLP) [19], with the benefit of avoiding pitch estimation required for LTP. In contrast, high-order LP (HOLP) [20] with the filter order covering T, without any imposed constraints will have many non-sparse prediction coefficients, where coefficients apart from the ones corresponding to STP and LTP might be seen as sub-optimal, i.e., increasing computational cost and having no or even negative contribution to prediction performance [5]. Similarly, conventional adaptive ANC algorithms, i.e., FXNLMS, act as an adaptive LP to find a causal filter when the causality constraint is violated [13, 21]. Hence, they might have the same problems as HOLP regarding the filter structure for voiced speech prediction, which could lead to limited performance for voiced speech reduction.

Since the sparsity measure corresponds to the so-called 0-norm [22], solutions to which cannot be found in polynomial time, the proposed block-based solutions to HOSpLP in [20, 22] used 1-norm regularisation criterion as a convex approximation to the 0-norm. There also exist different sparsity-aware adaptive algorithms broadly classified into two categories, namely, zero-attracting and the proportionate-type algorithms [23,24]. The last, particularly the improved proportionate NLMS (IPNLMS) algorithm, is a popular adaptive filter used to identify sparse systems in acoustic echo cancellation applications [4, 25, 26]. Also, the proportionate idea of the IPNLMS algorithm coincides with the idea of SLTPj, which is based on speech correlations, and as a result, is beneficial for our application when applied to solve the HOSpLP problem.

In this paper, we propose a fixed-filter FF ANC system for headphone applications, HOSpLP-ANC, which aims at attenuating voiced speech. It comprises an adaptive HOSpLP of speech ahead in time, thus overcoming the delay D, which violates the causality constraint. More specifically, we propose to apply an adaptive IPNLMS algorithm for the HOSpLP problem. As HOSpLP models SLTPj, it is expected that HOSpLP-ANC will outperform FXNLMS ANC and provide at least comparable performance to SLTPj-ANC, with the benefit of avoiding pitch estimation and reducing computational complexity. This is a clear advantage, especially in multi-speaker environments, e.g., open offices, since multi-pitch estimation is still a complicated problem [27, 28]. Note that the proposed HOSpLP-ANC is designed to improve voiced speech attenuation. It is intended to work alongside with, e.g., the conventional ANC system to attenuate other types of noise, including, e.g., unvoiced speech. However, this paper focuses only on the HOSpLP-ANC.

The paper is organised as follows. HOSpLP-ANC with residual error analysis is described in Section 2. Section 3 presents the idea of HOSpLP and the IPNLMS algorithm to solve it. Simulation results of voiced speech attenuation in the single and multiple speaker scenario are presented in Section 4. Section 5 concludes the paper.

2. PROPOSED FIXED-FILTER ANC SYSTEM

The proposed HOSpLP-ANC system, shown in Fig. 2, is for a single ear cup, which uses one reference and one error microphone to measure the incoming disturbance—speech $x(n)$ and the residual error $e(n)$. Two ear cups are independent in signal processing and ANC. The optimal transfer function of the FF ANC filter is

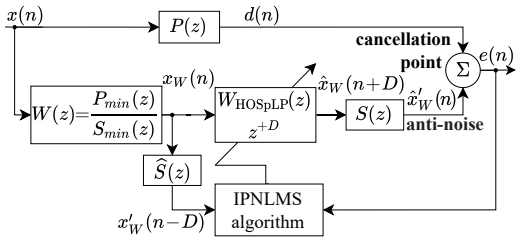

Figure 2: Simplified diagram of fixed-filter FF HOSpLP-ANC. The adaptive $W_{\text{HOSpLP}}(z)$ compensates for the delay D in $S(z)$.

$W^{\text{o}}(z) = P(z)/S(z)$ [6]. Since $P(z)$ and $S(z)$ include acoustic propagation paths and $S(z)$ also has the latency of an ANC processing unit, they are non-minimum phase and can be expressed as $P(z)$=$P_{min}(z)z^{-AD_P}$ and $S(z)$=$S_{min}(z)z^{-(ED_S+AD_S)}$, where $(\cdot)_{min}$ denotes the minimum-phase part. A causal $W^{\text{o}}(z)$ can only be realized if $P(z)$ contain a delay of at least equal length as $S(z)$.

In the HOSpLP-ANC, the causal FF fixed-filter $W(z)$ is calculated by taking the minimum-phase part of $P(z)$ and $S(z)$, i.e., $W(z)$=$P_{min}(z)/S_{min}(z)$. The delay part z^{-D} is compensated by the proposed $W_{\text{HOSpLP}}(z)$, which predicts $x_W(n)$ D samples ahead in time, resulting in $d(n)$ and $\hat{x}'_W(n)$ being aligned in time at the cancellation point, with the residual error

$$\begin{aligned} E(z) &= [P(z)X(z) - W(z)\hat{X}(z)z^{+D}S(z)] = [P(z)X(z) \\ &\quad - P_{min}(z)\hat{X}(z)z^{+(ED_S+AD_S-AD_P)}z^{-(ED_S+AD_S)}] \\ &= [X(z) - \hat{X}(z)]P(z). \end{aligned} \quad (1)$$

In this case, the performance of the proposed HOSpLP-ANC system will depend on the accuracy (in terms of phase and amplitude) of the predicted signal $\hat{X}(z)$ compared to the original signal $X(z)$, i.e., connected to the prediction performance. Without compensation for z^{-D}, i.e., bypassing $W_{\text{HOSpLP}}(z)$, it will lead to a significant decrease in the ANC performance [10–14], with the residual error given as $E_D(z) = [P(z) - W(z)S(z)]X(z) = [1 - z^{-D}]P(z)X(z)$.

The work here is focused on the prediction, which compensates for the delay D. Therefore, other challenges inherent in fixed-filter ANC design, e.g., the changes in $P(z)$, $S(z)$ and $\widehat{S}(z)$ due to the physiology of the ear and their influence on the performance, are not considered and are beyond the scope of the paper. The performance of FF ANC system depends on coherence between the reference and the error microphone [6]. For high ANC performance, it is necessary to have very high coherence [6]. This can be achieved by placing the reference microphone close to the dominant noise path, e.g., a rear vent, which will expand the frequency range where the coherence is high and will make $P(z)$ independent of the speaker location. In this paper, we consider such a design of the headphones.

3. HIGH-ORDER SPARSE LINEAR PREDICTION

We consider the following LP model, where a speech sample $x(n)$ is written as a linear combination of old M samples [21, 22]:

$$x(n) = \mathbf{a}^T \mathbf{x}(n - D) + r(n), \quad (2)$$

where $\mathbf{a} = [a_1, \ldots, a_M]^T$ is a vector with prediction coefficients, $\mathbf{x}(n-D) = [x(n-D), \ldots, x(n-D-M+1)]^T$ is a vector of M old speech samples and $r(n)$ is the prediction error. When $D > 1$, (2) becomes multi-step LP, i.e., predicting D samples ahead in time.

HOSpLP was presented in [19], where the approach was to impose sparsity into the HOLP coefficients by adding a regularisation criterion with a 1-norm to vector $\mathbf{a}$ while retaining a 2-norm minimisation criterion on the prediction error. The obtained convex optimisation problem was solved with iterative interior-point block-based methods [22] and other fast iterative block-based methods [20].

We propose to solve the HOSpLP problem by applying an adaptive IPNLMS algorithm based on the same 2-norm minimisation criterion on the prediction error, i.e., minimising the mean square error:

$$\mathbb{E}\{|r(n)|^2\} = \mathbb{E}\{|x(n) - \mathbf{a}^T\mathbf{x}(n-D)|^2\}. \quad (3)$$

However, what helps to create the sparse structure of $\mathbf{a}$, i.e., to obtain a solution corresponding to HOSpLP, is the proportionate idea of the IPNLMS algorithm. The idea is to update each filter coefficient independently of the others by adjusting the adaptation step-size in proportion to the magnitude of the estimated filter coefficient. In this way, larger coefficients are emphasized and receive a larger step-size, thus increasing the convergence rate of that coefficient [25, 26]. In the view of HOSpLP, the proportionate idea of the IPNLMS algorithm is beneficial for our application, such that filter coefficients with a large magnitude will correspond to the short- and long-term correlations of voiced speech, i.e., STP and LTP. The IPNLMS algorithm is summarised by the following equations [25]:

$$r(n) = x(n) - \mathbf{a}^T(n-1)\mathbf{x}(n-D), \quad (4)$$

$$\mathbf{a}(n) = \mathbf{a}(n-1) + \frac{\mu\mathbf{K}(n-1)\mathbf{x}(n-D)r(n)}{\mathbf{x}^T(n-D)\mathbf{K}(n-1)\mathbf{x}(n-D)+\delta}, \quad (5)$$

where μ is the normalized step-size and δ is the regularization. The diagonal matrix $\mathbf{K}(n-1) = \mathrm{diag}([k_1(n-1), \ldots, k_M(n-1)])$ adjusts the step-sizes of the individual prediction coefficient [25], with

$$k_m(n) = \frac{1-\alpha}{2M} + (1+\alpha)\frac{|a_m(n)|}{2\|\mathbf{a}(n)\|_1 + \varepsilon}, \quad (6)$$

where $-1 \leq \alpha < 1$, ε is a small positive number to avoid a division by zero. For $\alpha = -1$, the IPNLMS and NLMS algorithms are identical [26], which makes the switch between the algorithms quite easy and might be beneficial for real-time applications, e.g., transforming the system from improved voiced speech attenuation to the conventional FXNLMS ANC attenuating other types of noise. As can be seen in (6), the l_1 norm is used in the IPNLMS algorithm to exploit the sparseness of $\mathbf{a}$, with $\|\mathbf{a}\|_1 = \sum_{m=1}^{M} |a_m|$. The regularization δ for the IPNLMS algorithm is $\delta = \delta_{\mathrm{NLMS}}(1-\alpha)/(2M)$ [25]. The complexity of the IPNLMS algorithm is twice that of NLMS [26].

4. SIMULATION RESULTS

4.1. Simulation conditions

For the following simulations, $P(z)$ and $S(z)$ shown in Fig. 3 were measured on a Jabra headphone prototype with a head and torso simulator in an anechoic chamber with a point source representing the ambient noise to have high coherence in $P(z)$, with $S(z)$ excluding the ANC processing unit. Depending on the factors discussed in Section 1, e.g., due to the latency of the ANC processing unit, which is a very device-dependent parameter [15], the causality constraint might be violated, with the amount of D affected by those factors. Therefore, the ANC performance of the proposed system will be investigated as a function of the additional delay D in $S(z)$. In this regard, for the simulations, $AD_P = 0$, thereby $P(z) = P_{min}(z)$, while $S(z) = S_{min}(z)z^{-D}$ and $\widehat{S}(z) = \widehat{S}_{min}(z)z^{-D}$.

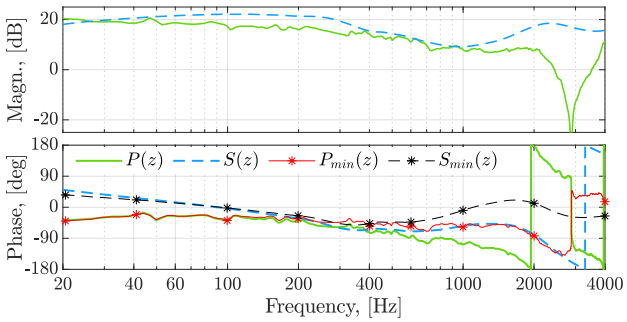

Figure 3: Measured $P(z)$, $S(z)$ and their minimum-phase parts.

The corresponding minimum-phase parts, shown in Fig. 3, were calculated with the real cepstrum method [29]. As an ambient noise input to the system, i.e., $x(n)$ in Fig. 2, we used speech examples from the TIMIT database [30] resampled at 8 kHz. The database contains clean speech signals from speakers with different characteristics (gender, age, pitch). In particular, we used 27, 28 and 20 seconds of concatenated speech signals for simulations with single, two and three speakers, respectively. The amount and possible combinations of male and female speech are balanced throughout the single and multi-speaker cases. We estimated performance on samples where at least one talker's speech was detected as voiced. For this, the voicing-unvoicing speech detection from [31] was used.

The impulse response length for $P(z)$, $S(z)$ and $\widehat{S}(z)$ is 100. The order of $W(z)$ is 128. It was trained on white Gaussian noise with the FXNLMS algorithm. Other prediction-based fixed-filter FF ANC systems used for comparison, namely, HOLP-ANC and SLTPj-ANC, based on the structure in Fig. 2 but using the relevant prediction scheme. HOLP-ANC is obtained by setting $\alpha = -1$ in (6). SLTPj-ANC was implemented using an adaptive NLMS algorithm, similar to [32]. The conventional adaptive FF ANC system, FXNLMS ANC [6], is also used for comparison. The simulation parameters listed below were found empirically. They determine the best possible performance of the systems under the given conditions. The order of HOSpLP, HOLP, and FXNLMS ANC is 100, which covers female and male pitch periods for the used test signals and in general [5]. The order of STP and LTP in SLTPj is 10 and 11, respectively. For all the systems $\delta_{\mathrm{NLMS}} = 10^{-3}$. For HOSpLP $\alpha = 0.85$, $\varepsilon = 10^{-2}$. Step-size μ for HOSpLP, HOLP, FXNLMS ANC is 0.17, for SLTPj it is 0.05. For SLTPj, T was estimated with the single pitch estimator from [31] every 10 ms on a 25 ms speech segment. For performance evaluation, the attenuation metric A, in (7), was calculated on voiced speech samples over a sliding window of 10 ms. The higher the A, the better the ANC speech attenuation.

$$A(n) = 10\log_{10}\left(\frac{\sum_{i=-I}^{I} d(n+i)^2}{\sum_{i=-I}^{I} e(n+i)^2}\right). \quad (7)$$

4.2. Results

1) Speech prediction. Fig. 4 shows an example of voiced speech prediction coefficients for two speakers, where HOLP has a non-sparse filter structure, while the proposed HOSpLP filter is sparse with nonzero coefficients corresponding to the short- and long-term correlations. The latter, e.g., can be seen in Fig. 4(b) at the female pitch period $T_{\mathrm{f}} = 29$ samples, male pitch period $T_{\mathrm{m}} = 67$ samples, and around integer multiples of T_{f}. The SLTPj filter is composed

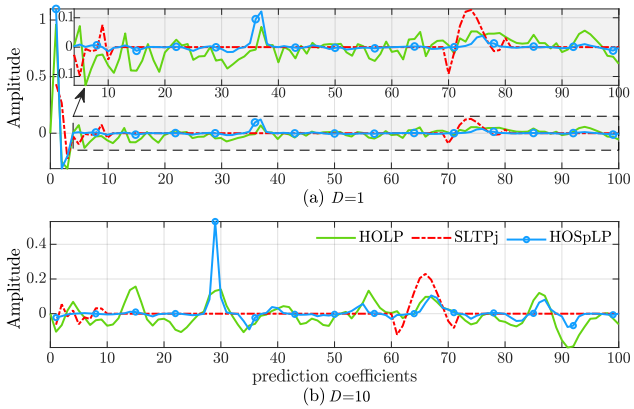

Figure 4: Example of the prediction coefficients for 2 speakers: female and male, and for the delay D of 1 (a) and 10 (b) samples.

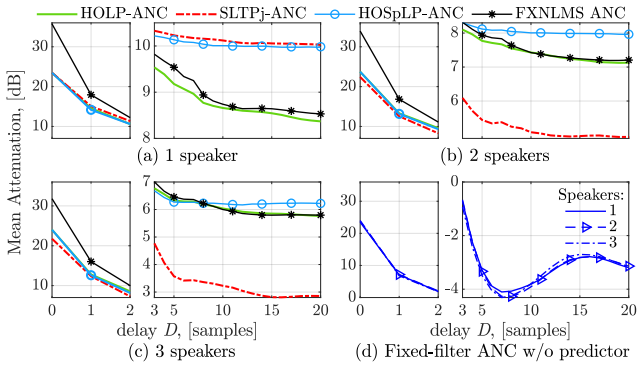

Figure 5: Mean attenuation of voiced speech for 1 (a), 2 (b) and 3 (c) speakers as a function of D for HOSpLP-ANC, other LP-based fixed-filter ANC, FXNLMS ANC and (d) ANC without predictor.

of the first 10 coefficients and 11 coefficients at the estimated pitch period. Since the used estimator from [31] is for the single pitch, the estimated T in SLTPj for the example in Fig. 4 corresponds to male speech. Comparing the first ten coefficients in Figs. 4(a) and 4(b), we can see that with increasing D, the short-term correlations are becoming weak, while the long-term are more pronounced.

2) Speech attenuation. Fig. 5 compares voiced speech attenuation performance of HOSpLP-ANC with different prediction-based fixed-filter ANC systems and the conventional adaptive FXNLMS ANC system. The causal ANC system, i.e., when $D=0$, shows quite high mean attenuation for voiced speech which is up to 24 dB and 35 dB for the fixed-filter and the adaptive FXNLMS ANC systems, respectively. Since the latter takes into account the input signal when modelling and updating the ANC filter, it provides higher attenuation. With a delay of 1 sample, the performance drops significantly for all the cases, and without compensating for D in fixed-filter ANC, as seen in Fig. 5(d), reaching even negative mean attenuation at $D=3$, meaning speech amplification. Integrating a predictor in the fixed-filter ANC allows to compensate for the delay D with the attenuation performance depending on the type of LP.

For the single-speaker case in Fig. 5(a), the proposed HOSpLP-ANC provides almost the same attenuation as SLTPj-ANC while avoiding pitch estimation. Compared to HOLP-ANC and FXNLMS ANC, HOSpLP-ANC has higher attenuation for $D>2$, outperforming HOLP-ANC on average by 1.4 dB and up to 1.8 dB at high D. As can be noticed, FXNLMS ANC follows HOLP-ANC with slightly higher attenuation. This behaviour is expected since they are both based on the NLMS algorithm resulting in lower performance than HOSpLP-ANC based on the IPNLMS algorithm. From the power spectral density (PSD) plot of $e(n)$ for $D=10$ samples in Fig. 6, we can see that, on average, all ANC systems are effective for voiced speech attenuation with delay compensation up to 1.5 kHz. HOSpLP-ANC outperforms FXNLMS ANC mostly at the frequency range of 650-1000 Hz for the given example in Fig. 6. At higher frequencies, i.e., around 2.7 kHz, FXNLMS ANC and SLTPj-ANC amplify speech slightly rather than attenuate it.

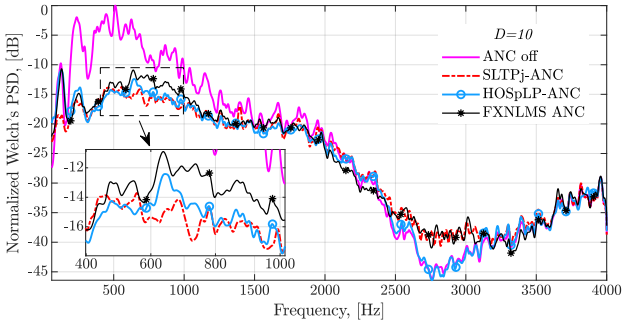

Figure 6: PSD of $e(n)$ for 1-speaker voiced speech from different ANC systems compensating for the delay D of 10 samples.

For the multi-speaker case in Figs. 5(b) and 5(c), the performance is lower for all the systems. As can be seen, HOSpLP-ANC still provides higher voiced speech attenuation than HOLP-ANC and FXNLMS ANC for two speakers and $D>3$, but the performance difference becomes less pronounced, especially for the three-speaker case, where HOSpLP-ANC has slightly higher attenuation for $D\geq 10$ and comparable performance to HOLP-ANC and FXNLMS ANC for $2<D<10$. This might be explained by the more complex and less sparse structure of the required prediction filter to be. Note that since SLTPj here is based on the single-pitch estimator, it has a relatively low but non-zero performance for the multi-speaker case. Multi-pitch estimation is still a complicated and challenging task [27, 28], requiring computationally demanding algorithms which are unsuitable for current headphones. In this regard, the advantage of HOSpLP avoiding pitch estimation can be of particular significance in the multi-speaker case.

5. CONCLUSION

We proposed a fixed-filter FF ANC system for headphone applications, HOSpLP-ANC, which aims at attenuating voiced speech and comprises high-order sparse linear prediction, exploiting the adaptive IPNLMS algorithm, to overcome the delay D which violates the causality constraint. Notably, HOSpLP provides high prediction performance of voiced speech by modelling the SLTPj scheme while avoiding pitch estimation. This can be of particular significance in the case of the complicated multi-pitch estimation scenario. Simulations show that HOSpLP-ANC outperforms conventional adaptive FF FXNLMS ANC in single and multi-speaker environments at a wide range of D, which is $3\leq D\leq 20$ samples (0.375 to 2.5 ms) and $4\leq D\leq 20$ samples (0.5 to 2.5 ms), respectively, at a sampling frequency of 8 kHz. Future work should focus on conducting subjective tests and experiments on a prototype.

6. REFERENCES

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

[2] 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, 2012.

[3] L. Lu, K. L. Yin, R. C. de Lamare, Z. Zheng, Y. Yu, X. Yang, and B. Chen, "A survey on active noise control in the past decade—Part I: Linear systems," *Signal Processing*, vol. 183, pp. 108039, 2021.

[4] L. Lu, K. L. Yin, R. C. de Lamare, Z. Zheng, Y. Yu, X. Yang, and B. Chen, "A survey on active noise control in the past decade–Part II: Nonlinear systems," *Signal Processing*, vol. 181, pp. 107929, 2021.

[5] W.-C. Chu, *Speech coding algorithms: foundation and evolution of standardized coders*, J. Wiley, New York, 2003.

[6] S. M. Kuo and D. R. Morgan, "Active noise control: a tutorial review," *Proc. IEEE*, vol. 87, no. 6, pp. 943–973, 1999.

[7] S. M. Kuo, S. Mitra, and W.-S. Gan, "Active noise control system for headphone applications," *IEEE Trans. Control Syst. Technol.*, vol. 14, no. 2, pp. 331–335, 2006.

[8] X. Shen, W.-S. Gan, and D. Shi, "Alternative switching hybrid ANC," *Applied Acoustics*, vol. 173, 2021.

[9] J. Fabry and P. Jax, "Primary path estimator based on individual secondary path for ANC headphones," in *Proc. IEEE ICASSP*, 2020, pp. 456–460.

[10] B. Rafaely, "Active noise reducing headset–an overview," in *Proc. INTER-NOISE*, 2001, vol. 2001, pp. 2144–2153.

[11] L. Zhang and X. Qiu, "Causality study on a feedforward active noise control headset with different noise coming directions in free field," *Applied Acoustics*, vol. 80, pp. 36–44, 2014.

[12] S. D. Snyder and C. H. Hansen, "The influence of transducer transfer functions and acoustic time delays on the implementation of the LMS algorithm in active noise control systems," *Journal of Sound and Vibration*, vol. 141, no. 3, pp. 409–424, 1990.

[13] K. Xuan and S. M. Kuo, "Study of causality constraint on feedforward active noise control systems," *Proc. IEEE ISCAS*, vol. 46, no. 2, pp. 183–186, 1999.

[14] M.-R. Bai, W. Pan, and H. Chen, "Active feedforward noise control and signal tracking of headsets: Electroacoustic analysis and system implementation," *J. Acoust. Soc. Am.*, vol. 143, no. 3, pp. 1613–1622, 2018.

[15] S. Liebich, J. Fabry, P. Jax, and P. Vary, "Signal processing challenges for active noise cancellation headphones," in *Speech Communication; 13th ITG-Symposium*, 2018.

[16] J. Wang, J. Zhang, J. Xu, C. Zheng, and X. Li, "An optimization framework for designing robust cascade biquad feedback controllers on active noise cancellation headphones," *Applied Acoustics*, vol. 179, 2021.

[17] B. Kovacevic, M. M. Milosavljevic, M. Veinović, and M. Marković, *Robust Digital Processing of Speech Signals*, Springer International Publishing AG, Cham, 2017.

[18] R. P. Ramachandran and P. Kabal, "Joint optimization of linear predictors in speech," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 5, pp. 642–650, 1989.

[19] D. Giacobello, M. G. Christensen, M. Murthi, S. H. Jensen, and M. Moonen, "Joint estimation of short-term and long-term predictors in speech coders," in *Proc. IEEE ICASSP*, 2009, pp. 4109–4112.

[20] T. L. Jensen, D. Giacobello, T. van Waterschoot, and M. G. Christensen, "Fast algorithms for high-order sparse linear prediction with applications to speech processing," *Speech Communication*, vol. 76, pp. 143–156, 2016.

[21] P. Diniz, *Adaptive Filtering: Algorithms and Practical Implementation*, Springer Nature Switzerland AG, Cham, 2020.

[22] D. Giacobello, M. G. Christensen, M. Murthi, S. H. Jensen, and M. Moonen, "Sparse linear prediction and its applications to speech processing," *IEEE Audio, Speech, Language Process.*, vol. 20, no. 5, pp. 1644–1657, 2012.

[23] S. S. Bhattacharjee, D. Ray, and Nithin V. George, "Adaptive modified Versoria zero attraction least mean square algorithms," *IEEE Trans. Circuits Syst. II*, vol. 67, no. 12, pp. 3602–3606, 2020.

[24] R. L. Das and M. Chakraborty, "Improving the performance of the pnlms algorithm using l_1 norm regularization," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 24, no. 7, pp. 1280–1290, 2016.

[25] J. Benesty and S. L. Gay, "An improved PNLMS algorithm," in *Proc. IEEE ICASSP*, 2002, vol. 2, pp. 1881–1884.

[26] C. Paleologu, J. Benesty, and S. Ciochina, "Sparse adaptive filters for echo cancellation," *Synthesis Lectures on Speech and Audio Processing*, vol. 6, no. 1, pp. 1–124, 2010.

[27] M. Wohlmayr, M. Stark, and F. Pernkopf, "A probabilistic interaction model for multipitch tracking with factorial hidden markov models," *IEEE Trans. Audio, Speech, and Language Process.*, vol. 19, no. 4, pp. 799–810, 2011.

[28] X. Li, Y. Sun, X. Wu, and J. Chen, "Multi-speaker pitch tracking via embodied self-supervised learning," in *Proc. IEEE ICASSP*, 2022, pp. 8257–8261.

[29] DSP Committee, *Programs for digital signal processing*, IEEE ASSP. IEEE Press, New York, 1979.

[30] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, and D. S. Pallett, "DARPA TIMIT acoustic-phonetic continous speech corpus CD-ROM. NIST speech disc 1-1.1," Tech. Rep., NASA STI/Recon, 1993.

[31] L. Shi, J. K. Nielsen, J. R. Jensen, M. A. Little, and M. G. Christensen, "Robust Bayesian pitch tracking based on the harmonic model," *IEEE Audio, Speech, Language Process.*, vol. 27, no. 11, pp. 1737–1751, 2019.

[32] Y. Iotov, S. M. Nørholm, V. Belyi, M. Dyrholm, and M. G. Christensen, "Computationally efficient fixed-filter ANC for speech based on long-term prediction for headphone applications," in *Proc. IEEE ICASSP*, 2022, pp. 761–765.