Non-Stationary Prediction for Addressing the Non-Causality Problem in Fixed-Filter ANC Headphones for 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 situations, 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 when the non-causality problem arises in such systems, which requires the optimal filters to be non-causal. The non-causality problem is due to the delay incurred by, e.g., digital processing or acoustic propagation paths. We propose a new fixed-filter feedforward ANC system, HCMP-ANC, which aims at attenuating speech in, e.g., office environments. Notably, it comprises a non-stationary harmonic chirp model-based prediction of speech ahead in time, thus overcoming the aforementioned delay. The results show that HCMP-ANC can outperform conventional adaptive feedforward ANC, for delays in the order of tens of samples at a sampling frequency of 8 kHz. By accounting for speech non-stationarity, HCMP-ANC can attenuate female speech in a wider frequency range of up to 3 kHz, while the conventional ANC is limited to 1.5 kHz.**

***Index Terms*—fixed-filter ANC for speech attenuation, causality, speech prediction, harmonic chirp model, ANC headphones.**

I. Introduction

Active noise control (ANC) technology has shown its effectiveness in attenuating different types of noise in many applications [1]–[3]. With the modern lifestyle, ANC is becoming an essential feature in headphones, headsets and small wireless earbuds with growing interest in both consumer and enterprise market. Among the different types of noise we deal with in everyday life, speech can be a very disturbing source of ambient noise, e.g., in crowded public spaces or open offices, reducing concentration and productivity. Therefore, it is increasingly important that ANC headphones also attenuate human voices effectively. However, speech attenuation by ANC headphones can be quite limited due to the non-stationarity of speech.

In ANC, an anti-noise signal with the same amplitude but the opposite phase is generated through a secondary source (e.g., headphone loudspeaker), cancelling unwanted noise at the desired cancellation point (e.g. 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 [1]–[3]. However, modern ANC headphones are typically based on either fixed feedforward (FF) or feedback (FB) filter or a combination of both (hybrid ANC) [4].

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

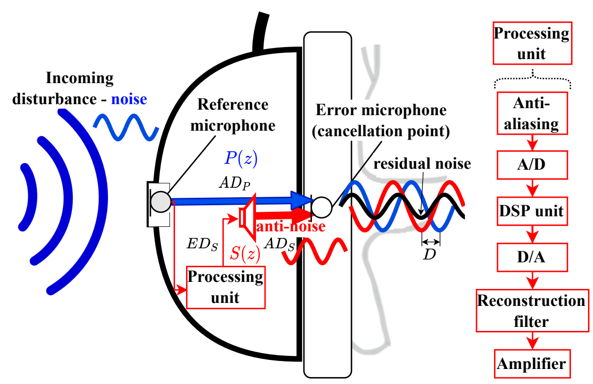

Fig. 1. Simplified modelling block diagram of FF ANC headphones with the extra delay $D = ED_S + AD_S - AD_P$ in $S(z)$.

Many factors affect ANC performance [1], [2], [4]–[10], one of those factors being the causality constraint. For FF ANC headphones, as shown in Fig. 1, the causality constraint is violated when the propagation delay AD_P of the disturbance between the reference and the error microphone of the primary path $P(z)$ is less than the electric, ED_S, and acoustic, AD_S, propagation delays of the anti-noise signal in the secondary path $S(z)$, i.e., $AD_P < ED_S + AD_S$, with the extra delay in $S(z)$, $D = ED_S + AD_S - AD_P$ [7]. When the causality constraint is violated, a prediction problem arises to compensate for the delay D [5]. The ANC causality has been investigated before [5]–[10]. However, the causality of ANC systems cancelling speech was not considered.

Due to the non-stationarity of speech, it requires the application of dedicated adaptive prediction schemes for higher prediction performance when compensating for D [11], [12]. A common approach for speech prediction is linear prediction (LP) [11]. The idea of LP is that a speech sample can be approximated as a linear combination of past samples. However, only the most recent 10-12 speech samples (at a sampling frequency, $f_s = 8$ kHz) and those at the pitch period of speech, T, contribute to the prediction performance [11]. This corresponds to short-term LP (STP) and long-term LP (LTP), the joint modeling of which, namely SLTPj, was proposed in [12]. In this regard, other speech samples than the ones corresponding to STP and LTP might be seen as suboptimal, i.e., increasing computational cost and having no or even negative contribution to prediction performance [11]. When the causality constraint is violated, conventional ANC algorithms, i.e., FXLMS or FXNLMS, act as an adaptive LP to find a causal filter [7], [13]. Depending on $P(z)$ and $S(z)$, it might have a filter order which is not optimal for speech prediction. Therefore, such ANC systems may have limited performance for speech reduction when compensating for D.

Apart from LP, sinusoidal (harmonic) modelling has been successfully applied to a broad range of speech processing problems [14]–[16]. In [17] it was shown that with uniform sampling, the harmonic model (HM) can be uniquely expressed using LP. In these models, speech is considered stationary during short analysis time intervals. However, it is well known that this assumption of stationarity does not hold [18], [19], which limits the accuracy of the models and increases prediction error. To account for speech non-stationarity, the harmonic-chirp model (HCM) was introduced, where a chirp parameter allows the frequency of the harmonics to change linearly within each segment [18], [19]. Moreover, higher speech harmonics are subject to quick changes, i.e., have higher temporal modulations [20]. In this regard, HCM should increase the prediction frequency range towards higher frequencies, i.e., better predict higher harmonics of speech. This advantage of HCM over LP and HM makes it a good candidate for addressing the prediction problem in ANC since the need for prediction is greater at higher frequencies as a constant time delay constitutes a larger phase difference at a higher frequency. Moreover, the human ear is most sensitive to sounds around 2–4 kHz [20].

In this paper, we propose a new fixed-filter FF ANC for headphones application, HCMP-ANC, comprising a non-stationary HCM-based prediction (HCMP) of speech ahead in time, thus overcoming the delay which creates the non-causality problem. Since HCM models non-stationary speech better, it is expected that HCMP-ANC will improve speech attenuation compared to conventional adaptive ANC, especially at higher frequencies. In this study, we consider the voiced part of speech since it is the main constituent of speech and normally has higher power than unvoiced speech [19]. Also, the voiced speech has much higher predictability than the unvoiced—stochastic part, which is almost unpredictable [11].

The paper is organised as follows. First, HCMP-ANC with residual error analysis is described in Section II. Section III presents the signal model and the proposed HCMP. Speech attenuation performance of the proposed HCMP-ANC compared to the conventional FXNLMS ANC and to SLTPj-ANC are presented in Section IV. Section V concludes the paper.

II. Proposed fixed-filter ANC system

The proposed HCMP-ANC system is depicted in Fig. 2. In most commercially available FF ANC headphones, at least one microphone in each ear cup, a reference microphone, is used to measure the incoming noise $x(n)$ [4]. In the FF ANC

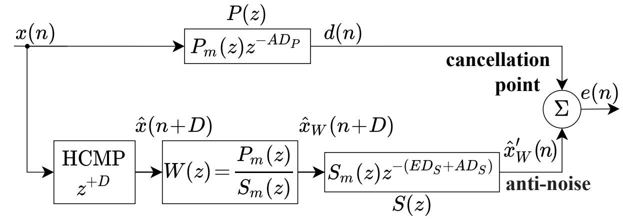

Fig. 2. Simplified diagram of the proposed fixed-filter FF ANC, HCMP-ANC. Adaptive HCMP compensates for the delay $D=ED_S+AD_S-AD_P$; $(\cdot)_m$ denotes the minimum-phase part of $P(z)$ and $S(z)$.

system, the goal is to match the internal disturbance signal $d(n) = x(n) * p(n)$ as accurately as possible in amplitude and inverted phase. Since $P(z)$ and $S(z)$ include acoustic propagation paths and $S(z)$ also has the latency of an ANC chip (i.e., the processing unit), they are non-minimum phase. Therefore, the causal FF fixed-filter $W(z)$ can be calculated by considering the minimum-phase part of $P(z)$ and $S(z)$. The delay part z^{-D} is compensated by the proposed HCMP which predicts $x(n)$ D samples ahead in time, resulting in both signals $d(n)$ and $\hat{x}'_W(n)$ being aligned in time at the cancellation point, with the residual error

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

As can be seen from (1), the performance of the proposed HCMP-ANC system depends on the accuracy (in terms of waveform approximation) of the predicted signal $\hat{X}(z)$ compared to the original signal $X(z)$. Without compensation for z^{-D}, the ANC performance will be significantly decreased [5]–[8]. The residual error, in this case, can be expressed as

$$E_D(z) = [P(z) - W(z)S(z)]X(z) = [1 - z^{-D}]P(z)X(z). \quad (2)$$

Depending on the ANC headphones design, the causality constraint might be violated due to a geometry problem. This is related to the small size of the ear cups (earbuds), which results in a small AD_P compared to AD_S combined with the processing (electric) latency ED_S. In addition, the amount of ED_S depends on the ANC chip and its algorithmic design [8]–[10]. The delay D might also be affected by the direction of the incoming noise [6] and improper headphone fit on the ear. The work here is focused on addressing the prediction problem. Therefore, other challenges inherent in fixed-filter ANC design, e.g., the changes in $P(z)$ and $S(z)$ due to the physiology of the ear and their influence on the ANC performance, are not considered and are beyond the scope of the paper. This allows for breaking down the complexity of an ANC system and focusing on the non-causality problem.

III. Proposed harmonic chirp model prediction

A single-pitch real-valued speech signal is modelled by $x(n) = s(n) + u(n)$, where $s(n)$ is the harmonic part of the signal, i.e., voiced speech, and $u(n)$ is the stochastic part, i.e., unvoiced speech, background noise, etc.

The real HCM for a voiced speech signal is given by [19]

$$s_{\text{HCM}}(n) = \sum_{l=-L}^{L} A_l e^{j\varphi_l(n)}, \tag{3}$$

where L is the number of harmonics, A_l is a real, non-zero amplitude of the l'th harmonic, with $A_0 = 0$, and $\varphi_l(n)$ is the corresponding instantaneous phase. In HCM, the instantaneous frequency ω_l is not stationary but changes linearly within a segment, i.e., $\omega_l(n) = l(\omega_0 + kn)$. The instantaneous phase $\varphi_l(n)$ is then found by integrating ω_l as [19]:

$$\varphi_l(n) = l\left(\omega_0 n + \frac{1}{2}kn^2\right) + \phi_l, \tag{4}$$

where $\omega_0 = 2\pi f_0/f_{\text{s}}$ is the normalized fundamental frequency, k is the normalized fundamental chirp rate and $\phi_l \in [0, 2\pi]$ is the initial phase of the l'th harmonic. The model in (3) can then be rewritten as [19]:

$$s_{\text{HCM}}(n) = \sum_{l=-L}^{L} \alpha_l e^{jl\left(\omega_0 n + k/2n^2\right)}, \tag{5}$$

where $\alpha_l = A_l e^{j\phi_l}$ is the complex amplitude of the l'th harmonic. A special case of HCM for $k = 0$ is the traditional HM, where ω_l is stationary within a segment, i.e., $\omega_l(n) = l\omega_0$, which gives [19]:

$$s_{\text{HM}}(n) = \sum_{l=-L}^{L} \alpha_l e^{jl\omega_0 n}. \tag{6}$$

As was demonstrated in [17], HM in (6) can be uniquely expressed as a linear combination of its previous $2L$ samples:

$$s_{\text{HM}}(n) = -\sum_{i=1}^{2L} a_i s_{\text{HM}}(n-i), \tag{7}$$

with $\sum_{i=0}^{2L} a_i \exp(-j\omega_l i) = 0$ relating $\{\omega_l\}$ to the LP coefficients, $\{a_i\}$, where $a_0 = 1$, $a_i = a_{2L-i}$. The relation in (7) holds for a stationary signal, i.e., a number of sinusoids [17]. Since ω_0 of speech is a non-linear parameter, it makes HM a non-linear problem for speech [21] and could lead to higher estimation and prediction errors. With application to speech and assuming its stationarity during short time intervals, (7) shows that using LP for speech prediction is more practical, and it does not make sense to do prediction with HM. Moreover, with some LP schemes, the estimation of ω_0 can be avoided [11]. However, speech is a non-stationary signal, even within short time intervals [18], [19], which is not taken into account by the LP model. Therefore, it makes sense to apply HCM for speech prediction. To the best of our knowledge, an LP equivalent of HCM, which could take into account speech non-stationarity, is not known.

In this paper, we propose HCM-based prediction (HCMP), which is done by extending the model in (5) in discrete-time, so predicted D samples of voiced speech ahead in time are:

$$\hat{s}_{\text{HCMP}}(n+D) = \sum_{l=-L}^{L} \widehat{\alpha}_l e^{jl\left(\widehat{\omega}_0(n+D) + \widehat{k}/2(n+D)^2\right)}, \tag{8}$$

and similar for the HM-based prediction (HMP), when $\widehat{k} = 0$ in (8), where $(\widehat{\cdot})$ denotes estimates of the corresponding model parameters that minimize the mean square error

$$\mathbb{E}\{|e_M(n)|^2\} = \mathbb{E}\{|x(n) - \widehat{s}(n)|^2\}. \tag{9}$$

Given that $\widehat{\omega}_0$ and $\widehat{k}$ are non-linear parameters they can be found using the non-linear least-squares (NLS) method minimizing (9), as it has been proved to be the most accurate and robust [21], [22]. A Kalman filter, similar to [23], [24], can also be used to update $\widehat{\omega}_0$ and $\widehat{k}$. When estimating $\widehat{\omega}_0$ and $\widehat{k}$ there is a trade-off between segment length and accuracy of the estimates [22]. A typical segment of 20-30 ms is used in this case. However, considering temporal modulations of speech [20], the optimal segment length when estimating $\widehat{\alpha}_l$ may be shorter to correctly capture speech formants. Therefore, once $\widehat{\omega}_0$ and $\widehat{k}$ are estimated, the amplitudes $\widehat{\alpha}_l$ can be found minimizing (9) by, e.g., either the least squares method with a recursive algorithm (RLS) or with LMS algorithm [13].

IV. Simulation results

A. Simulation conditions

For the following simulations, $P(z)$ and $S(z)$ were measured on a Jabra headphone prototype in an anechoic chamber with directional noise on a head and torso simulator, with $S(z)$ excluding the ANC processing unit. Depending on the factors discussed in Section II, the ANC system might become non-causal, with the amount of D depending on those factors. Therefore, the ANC performance of the proposed system will be investigated as a function of the delay D in $S(z)$. In this regard, for the simulations, $AD_P = 0$, thereby $P(z) = P_m(z)$, while $S(z) = S_m(z)z^{-D}$ and $\widehat{S}(z) = S(z)$ for FXNLMS ANC.

As an ambient noise input to the system, i.e., $x(n)$ in Fig. 2, we used 27 and 23 seconds of concatenated clean speech signals sampled at 8 kHz from the TIMIT [25] and the NOIZEUS [26] databases. The amount of female and male speech is equal in the selected material. The simulation parameters listed below were found empirically. They determine the best possible performance of the systems under the given conditions. For HCMP, $\widehat{\omega}_0$ and $\widehat{k}$ were estimated using the NLS estimator from [21] with the segment length of 161 samples with symmetric time indices [21]. The maximum model order L for estimation is 15. For prediction, L was set to have all harmonics up to $f_{\text{s}}/2$. The amplitudes $\widehat{\alpha}_l$ were estimated with the least squares method over a window of length 63 and 119 for female and male speech, respectively. The smaller optimal window length for female speech can be explained by higher temporal modulations, i.e., faster changes in the amplitude of the female speech harmonics [20]. For SLTPj and HMP, T and $\widehat{\omega}_0$ were estimated with the pitch estimator from [21] with the same settings as above. SLTPj was implemented using an adaptive NLMS algorithm as in [27] with the order of 10 and 5 for STP and LTP parts, respectively. The order of FXNLMS ANC is set to 72, which also covers female and male pitch periods for the used test signals. The order of $W(z)$ is 72. For all the systems regularisation $\delta_{\text{NLMS}} = 10^{-3}$, step-size μ

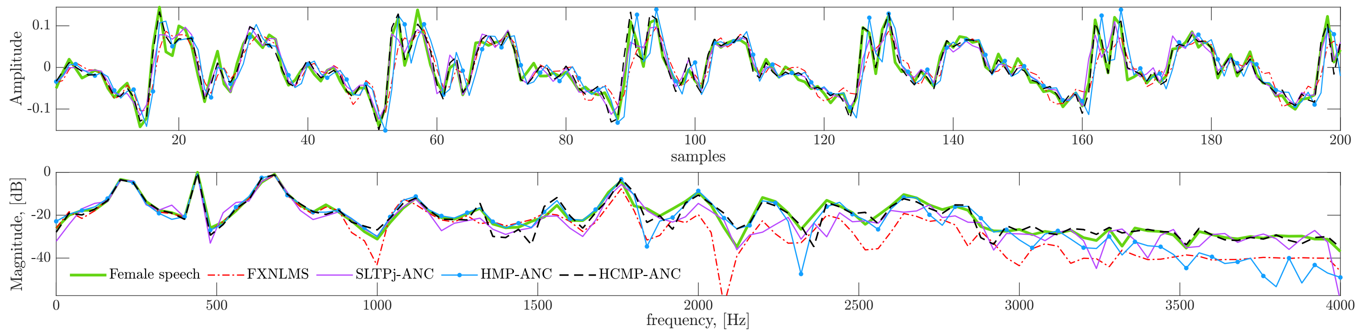

Fig. 3. Illustration of a non-stationary voiced segment of $x(n)$—female speech and its prediction $\hat{x}(n+D)$, $D=5$ samples. Time-domain and its spectra.

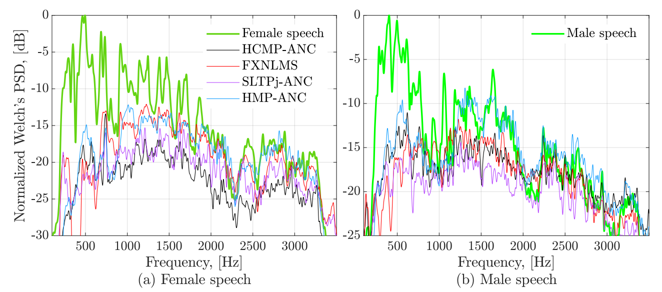

Fig. 4. PSD of $x(n)$—female (a), male (b) voiced speech and prediction error for $\hat{x}(n+D)$, $D=5$ samples. Averaged over all female and male signals.

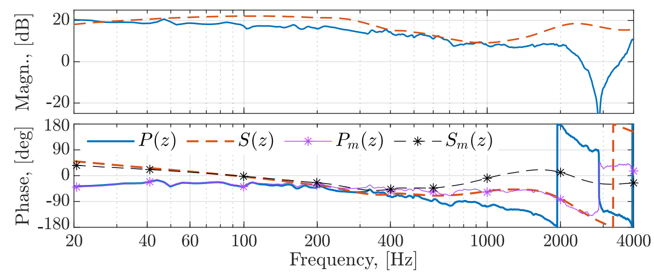

Fig. 5. Measured primary path $P(z)$, secondary path $S(z)$ and their minimum-phase parts calculated with the real cepstrum method [29].

is 0.15 for FXNLMS ANC and 0.16 for SLTPj. The estimation and prediction were done on a sample-by-sample basis. While the entire speech signals were used for estimation and prediction, only voiced samples were considered to evaluate performance. For this purpose, the voiced-unvoiced detection [28] was used with the same settings as above. To measure average ANC speech attenuation performance, the attenuation metric A, in dB, was used. The higher the attenuation, the better the ANC performance. The metric was calculated on voiced speech samples over a sliding window of 25 ms as:

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

B. Results

1) Simulation of $P(z)$ and $S(z)$ as pure delays, i.e., $\{P_m(z), S_m(z), W(z)\} = 1$. As the work here is focused on addressing the non-causality problem in ANC, first, we evaluate and compare the performance on the entire speech frequency range without the bias of $P(z)$ and $S(z)$ from a specific headphone design. The attenuation performance in this case will be equivalent to the prediction performance [7].

In Fig. 3 we show an example of a non-stationary segment of female speech, where HCMP-ANC generally has better performance and a quite noticeable performance improvement at higher frequencies. According to the power spectral density (PSD) plot in Fig. 4, the proposed system for female speech has a lower prediction error in a wide frequency range, especially at higher frequencies and up to 3 kHz, therefore, predicting higher frequencies better than other ANC systems. This might be explained by the ability of HCMP-ANC to account for speech non-stationarity, which can also be seen when comparing to HMP-ANC and shows its importance for speech. In contrast, FXNLMS ANC has quite a poor performance at frequencies above 1.5 kHz and generally higher prediction error, as seen in Figs. 3 and 4(a). Compared to HCMP-ANC, SLTPj-ANC has a higher error and still limited performance at higher frequencies for female speech, as seen in Fig. 4(a). However, it outperforms the proposed system for male speech, providing a lower prediction error, as shown in Fig. 4(b).

2) Measured $P(z)$ and $S(z)$ are shown in Fig. 5. The average speech attenuation performance when compensating for the delay D is shown in Fig. 6. When the system is causal, i.e., $D = 0$, it allows for an average speech attenuation of about 30 dB, which may not be achievable in practice due to other factors affecting ANC performance. However, with a delay of 1 sample, as seen in Fig. 6(c), the performance drops significantly, and without compensating for D, reaching even negative attenuation, meaning noise amplification. As seen in Fig. 6 (a), the proposed HCMP-ANC provides the highest average female speech attenuation for $D > 1$ among all considered ANC systems outperforming the conventional FXNLMS ANC and SLTPj-ANC.

Male speech attenuation is shown in Fig. 6(b). As in the case of female speech, HCMP-ANC outperforms FXNLMS ANC. However, SLTPj-ANC has a better performance than the proposed system. This might be explained by the fact that for male speech more model parameters, i.e., complex amplitudes $\widehat{\alpha}_l$, should be estimated, which can make HCMP-ANC and HMP-ANC more susceptible to estimation and prediction errors

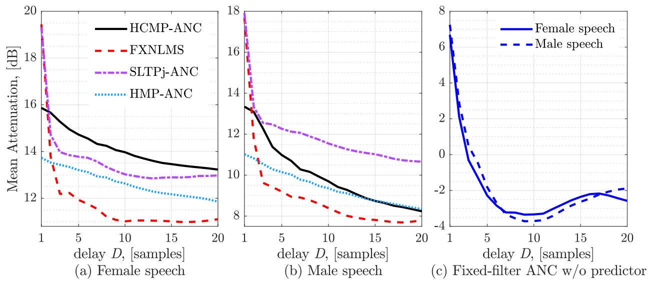

Fig. 6. ANC performance with measured $P(z)$ and $S(z)$: mean attenuation of female (a) and male (b) voiced speech as a function of D. Comparison of the proposed HCMP-ANC with the conventional FXNLMS ANC, other prediction-based fixed-filter ANC and (c) fixed-filter ANC without predictor.

than LP-based systems. This is because the lower fundamental frequency in male speech leads to more harmonics with their finer spacing [20]. Moreover, male speech produces higher spectral (frequency) modulations, varying over harmonics [20]. Therefore, a better solution would be a combination of HCMP-ANC for female speech and SLTPj-ANC for male speech. The decision, e.g., can be made based on the estimated fundamental frequency, which is anyway required for both systems.

V. Conclusion

We proposed a new fixed-filter FF ANC system for headphone applications, HCMP-ANC, which aims at attenuating speech and comprises a non-stationary harmonic chirp model-based prediction to overcome the delay creating the non-causality problem. Simulations show that HCMP-ANC is particularly good for female speech. It outperforms conventional adaptive FF FXNLMS ANC for female and male speech as well as linear prediction-based fixed-filter FF ANC system, SLTPj-ANC, for female speech at a wide range of the delay D, i.e., $1 < D \leq 20$ samples at $f_s = 8$ kHz. By accounting for speech non-stationarity, HCMP-ANC can attenuate female speech in a wider frequency range of up to 3 kHz, while FXNLMS ANC is limited to 1.5 kHz. However, for male speech, SLTPj-ANC appears to be a better solution. Therefore, a combination of the two methods would likely work best. Future work should focus on conducting subjective tests.

References

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

[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] J. Fabry and P. Jax, "Primary path estimator based on individual secondary path for ANC headphones," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 2020, pp. 456–460.

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

[6] 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.

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

[8] 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.

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

[10] 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.

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

[12] 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.

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

[14] R. McAulay and T. Quatieri, "Speech analysis/synthesis based on a sinusoidal representation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 34, no. 4, pp. 744–754, 1986.

[15] C. A. Rodbro, M. G. Christensen, S. V. Andersen, and S. H. Jensen, "Compressed domain packet loss concealment of sinusoidally coded speech," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 2003, vol. 1, pp. I–104.

[16] J. Lindblom, "A sinusoidal voice over packet coder tailored for the frame-erasure channel," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 5, pp. 787–798, 2005.

[17] Y. Chan, J. Lavoie, and J. Plant, "A parameter estimation approach to estimation of frequencies of sinusoids," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 29, no. 2, pp. 214–219, 1981.

[18] Y. Pantazis, O. Rosec, and Y. Stylianou, "Chirp rate estimation of speech based on a time-varying quasi-harmonic model," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 2009, pp. 3985–3988.

[19] S. M. Nørholm, J. R. Jensen, and M. G. Christensen, "Enhancement and noise statistics estimation for non-stationary voiced speech," *IEEE Trans. Audio, Speech, and Language Process.*, vol. 24, no. 4, pp. 645–658, 2016.

[20] J. Schnupp, I. Nelken, and A. King, *Auditory Neuroscience: Making Sense of Sound*, MIT Press, 2011.

[21] T. L. Jensen, J. K. Nielsen, J. R. Jensen, M. G. Christensen, and S. H. Jensen, "A fast algorithm for maximum-likelihood estimation of harmonic chirp parameters," *IEEE Trans. Signal Process.*, vol. 65, no. 19, pp. 5137–5152, 2017.

[22] J. K. Nielsen, T. L. Jensen, J. R. Jensen, M. G. Christensen, and S. H. Jensen, "Fast fundamental frequency estimation: Making a statistically efficient estimator computationally efficient," *Signal Processing*, vol. 135, pp. 188–197, 2017.

[23] M. G. Christensen, "A method for low-delay pitch tracking and smoothing," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 2012, pp. 345–348.

[24] L. Shi, J. K. Nielsen, J. R. Jensen, M. A. Little, and M. G. Christensen, "A Kalman-based fundamental frequency estimation algorithm," in *Proc. IEEE Workshop on Appl. of Signal Process. to Aud. and Acoust.*, 2017, pp. 314–318.

[25] 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," *NASA STI/Recon technical report n*, vol. 93, pp. 27403, 1993.

[26] Y. Hu and P. Loizou, "Subjective comparison and evaluation of speech enhancement algorithms," *Speech communication*, vol. 49, no. 7, pp. 588–601, 2007.

[27] 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 Int. Conf. Acoust., Speech, Signal Processing*, 2022, pp. 761–765.

[28] 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 Trans. Audio, Speech, and Language Process.*, vol. 27, no. 11, pp. 1737–1751, 2019.

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