COMPUTATIONALLY EFFICIENT FIXED-FILTER ANC FOR SPEECH BASED ON LONG-TERM PREDICTION FOR HEADPHONE APPLICATIONS

Yurii Iotov[*,1,2], *Sidsel Marie Nørholm*[2], *Valiantsin Belyi*[2], *Mads Dyrholm*[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, mdyrholm}@jabra.com

ABSTRACT

In some situations, such as open office spaces, speech can play the role of an unwanted and disturbing source of noise, and ANC headphones or earbuds might help to solve this problem. However, ANC in modern headphones is often based on a pre-calculated fixed-filter for practical reasons, like stability and cost. Moreover, in some cases the optimal filter is non-causal, which cannot be realized with such a filter, and ANC attenuation performance will be significantly decreased. In this paper we propose to solve the causality problem in feedforward fixed-filter ANC systems by integrating a long-term linear prediction filter to predict the incoming disturbance, here speech, by the same amount of samples ahead in time, as the non-causal delay. The proposed ANC system outperforms conventional adaptive feedforward ANC systems in terms of computational complexity, showing comparable or better results on voiced speech attenuation at non-causal delays from 4 to 18 samples (0.5 to 2.25 ms) at a sampling frequency of 8 kHz.

***Index Terms*—** Speech attenuation, Fixed-filter ANC, Long-term linear prediction, Causality, ANC headphones.

1. INTRODUCTION

Active noise control (ANC) systems have a rich history and many different applications [1–3]. Recently, with the general industrial and technological development, modern lifestyle, the noise pollution problem is becoming more pressing and significantly affects our daily life at different levels. In this regard, ANC systems are evermore in demand and widespread in the consumer market. Among the various noise sources we are dealing with in everyday life, speech has become an increasingly unwanted source of noise. For example, when working in an open office, when people are talking on the phone nearby, etc., and speech in these cases may be even more annoying than other types of noise.

ANC is a method for dealing with unwanted noise wherein noise with similar amplitude and opposite phase (i.e., an anti-noise signal) is emitted by a secondary source, e.g., headphone loudspeaker, and thereby cancels the noise at the desired cancellation point, e.g., at the eardrum. ANC systems can be adaptive, i.e., based on an adaptive filter (AF), or based on a fixed-filter, and configured in a feedforward (FF) structure, a feedback (FB) structure, or combined [2]. State-of-the-art ANC systems with AF (hereafter adaptive ANC) [2–7] mostly incorporate least mean squares (LMS) based algorithms, like Filtered-X LMS (FXLMS) or Filtered-X Normalized LMS (FXNLMS). Whereas ANC with a fixed-filter (hereafter fixed-filter ANC) is based on a pre-calculated filter. Due to real-time application challenges of adaptive ANC systems, including stability, computational complexity, and cost, fixed-filter ANC implementations are often preferable and used for modern ANC headphones [6, 8].

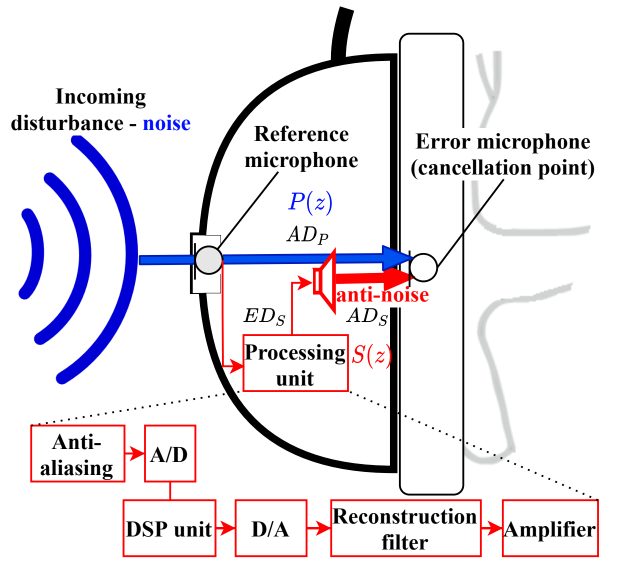

Fig. 1. Simplified modelling block diagram of FF ANC headphones.

Fig. 1 depicts a simplified block diagram of FF ANC headphones. The primary path $P(z)$ is the propagation path between the reference and the error microphone (cancellation point), while the secondary path, i.e., the anti-noise generation path $S(z)$, might include transfer functions of anti-aliasing filter, any A/D-D/A converters, DSP unit, reconstruction filter, power amplifier, loudspeaker and transfer function of acoustic propagation from the loudspeaker to the cancellation point. The performance of an ANC system depends on several factors, one of them being the causality constraint [1–3]. When the signal propagation delay AD_P of the primary path is less than the electric ED_S and acoustic AD_S propagation delays in the $S(z)$, i.e., $AD_P < ED_S + AD_S$, the causality constraint is violated, and the optimal ANC filter is non-causal [9]. The difference in propagation or the non-causal delay D is given by $D = ED_S + AD_S - AD_P$. The origin of D, by the example of ANC headphones, might be due to a geometry problem, related to the small size of headphones and short AD_P, combined with a contribution from ED_S due to delays in the components mentioned above. The non-causal delay might be caused also by different noise coming directions [9]. Additionally, improper headphone fit on the ear might lead to D as well.

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

When the causality constraint is violated, a prediction problem arises, and, in the case of adaptive ANC systems, the AF acts as a predictor also to find a causal filter [10]. The ANC performance, in this case, depends on the predictability of the noise to be cancelled (e.g., whether it is highly predictable tonal, narrow-band or unpredictable white noise), and might be quite poor. This was shown and discussed in several studies before, and causality in ANC systems has been a major topic within ANC during the past decades [9–13]. For a causal fixed-filter ANC system, the occurrence of D which was not or cannot be considered in the fixed-filter design stage will make the fixed-filter unable to compensate for this delay, and the performance of this ANC system will be significantly decreased [9–13]. Hence, it is critical to solve the prediction problem. Among numerous studies dedicated to noise reduction and ANC in general the case when the noise to be cancelled is speech has so far only been the subject of a few studies [14–17], where the implemented solutions are based on the same conventional adaptive ANC [2]. However, the causality of ANC systems cancelling speech was not considered in these studies.

The work presented here focuses on solving the prediction problem arising in a non-causal FF fixed-filter ANC system, when the noise to be cancelled is speech. It is known that speech can be predicted, and a common approach for this is linear prediction (LP), which has been studied comprehensively and used extensively in speech processing since the last century [18–20]. It was shown in [18] that, specifically, the long-term speech LP scheme is efficient in terms of prediction performance and computational complexity. Thus, we propose an integration of long-term LP into FF fixed-filter ANC system to solve the non-causal delay problem by predicting the incoming disturbance—speech D samples ahead in time. The prediction in this study will focus on voiced speech since it was demonstrated in [18] that voiced speech has much higher predictability than unvoiced speech due to the low correlation between unvoiced samples. Voiced speech is also the main constituent of speech and normally has higher power than unvoiced speech.

The paper is organized as follows: the proposed ANC system is described in Section 2. In Section 3 the joint short and long-term speech LP scheme is presented. The simulation results and computational complexity of speech prediction and speech attenuation of the proposed ANC system in comparison with the conventional adaptive ANC are presented in Section 4. Section 5 concludes the paper.

2. PROPOSED FIXED-FILTER ANC SYSTEM

We will now present the proposed FF fixed-filter ANC system with integrated predictor, an overview of which is depicted in Fig. 2. In the case of ANC headphones, only one microphone in each ear cup is typically used to measure the incoming noise $x(n)$.

In a conventional causal FF ANC system [2] the anti-noise signal, which is the ANC filter output convolved with $S(z)$, is ideally identical in amplitude but opposite in phase to the primary noise signal d at the cancellation point for perfect cancellation, resulting in a residual error of zero [2, 9], i.e.,

$$E(z) = [P(z) - W^{\text{o}}(z)S(z)]X(z) = 0. \quad (1)$$

This means that an ideal (optimal) transfer function of the FF ANC filter $W^{\text{o}}(z)$ is

$$W^{\text{o}}(z) = \frac{P(z)}{S(z)}. \quad (2)$$

So, the FF ANC filter has to model $P(z)$ and the inverse of $S(z)$, thus $W^{\text{o}}(z)$ can only be realized if $S(z)$ is a minimum phase and whereby a stable and causal inverse exists. However, perfect cancellation may not be realizable in practice [8, 9]. In the case of an FF fixed-filter ANC system based on a stable and causal filter, found based on (2), i.e., based on a-priori calculation using pre-measured $P(z)$ and $S(z)$, ANC performance of this system will depend to a great extent on the accuracy of the pre-measured $P(z)$ and $S(z)$, and their stationarity.

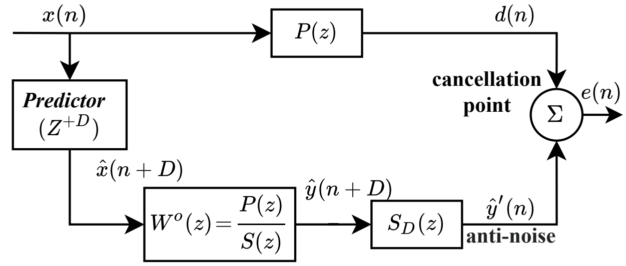

Fig. 2. Simplified block diagram of the proposed FF fixed-filter ANC system with predictor and non-causal delay D.

As a proof of concept and to simplify the analysis, we assume that the fixed-filter ANC system considered here is based on (2), and in the case of a non-causal delay D, the non-causal secondary path $S_D(z)$ consists of a minimum phase $S(z)$, and a pure delay z^{-D}, i.e., $S_D(z) = S(z)z^{-D}$. With the occurrence of D, the fixed-filter ANC system will be unable to take it into account, and the signals $d(n)$ and $y'(n)$ will not be aligned in time at the cancellation point, meaning that the residual error will be non-zero, i.e.,

$$E_D(z) = [P(z) - W^{\text{o}}(z)S(z)z^{-D}]X(z) = [1 - z^{-D}]P(z)X(z). \quad (3)$$

As a result, the ANC performance will in this case be significantly decreased [2, 9–12].

In the proposed FF fixed-filter ANC system the issue of a non-causal delay D is solved by integrating the long-term linear prediction filter into the ANC system, as depicted in Fig. 2. The purpose of the prediction is to predict D samples ahead in time, resulting in both signals $d(n)$ and $\hat{y}'(n)$ being aligned in time at the cancellation point, with the residual error being

$$\begin{aligned} E_{\text{proposed}}(z) &= [X(z)P(z) - \hat{X}(z)z^{+D}W^{\text{o}}(z)S(z)z^{-D}] \\ &= [X(z) - \hat{X}(z)]P(z). \end{aligned} \quad (4)$$

In this case, the ANC performance of the system depends on prediction performance, i.e., the accuracy of the predicted signal $\hat{X}(z)$ compared to the original signal $X(z)$.

3. MULTI-STEP PREDICTION OF SPEECH

The nature of speech is quite complex, and it tends to be highly non-stationary. Speech sounds can be broadly divided into voiced and unvoiced speech. Voiced speech is approximately periodic and exhibits a structure defined by a fundamental frequency. Unvoiced speech has a stochastic nature a continuous and smooth spectrum and is therefore almost unpredictable [18, 19]. For this reason, we are here focusing on voiced speech. The integrated predictor in the proposed ANC system in Fig. 2 is based on the LP model, the fundamental idea of which is that a speech sample can be approximated as a linear combination of past samples. The experiments reported in [18] show that the prediction order must be high enough to include at least one pitch period in order to model adequately the voiced speech. However, only the first 8-10 coefficients and the ones at the pitch period have a strong contribution to an increase of prediction gain. Therefore, the remaining coefficients can be eliminated, leading to a reduced computational complexity [18]. This scheme represents the idea of the long-term LP, where a short-term predictor (STP) is combined with a long-term linear predictor (LTP). Linear prediction typically considers prediction by only one sample ahead in time, but for our purpose, multi-step LP is required, predicting the D-th sample ahead in time. This can be obtained from the standard form of long-term linear prediction as [20]

$$\hat{x}(n+D)=\sum_{k=1}^{M} a_k x(n-k+1)+\sum_{k=-Q}^{Q} b_k x(n-T-k+D), \quad (5)$$

with M and Q being STP and LTP prediction orders, a_k and b_k are prediction coefficients, and T being the pitch period. There are two steps in the LP process: *estimation*, where a_k and b_k coefficients are found, and *prediction*, when the estimated coefficients are used to predict the future samples, i.e., $\hat{x}(n+D)$. Due to the dynamic nature of speech, estimation and prediction are here performed in an adaptive fashion, sample-by-sample. The coefficients a_k and b_k are found jointly by minimizing the estimation error

$$e_{\text{est}}(n)=x(n)-\sum_{k=1}^{M} a_k x(n-k-D+1)-\sum_{k=-Q}^{Q} b_k x(n-T-k), \quad (6)$$

in the mean square sense. This long-term linear prediction problem can be formulated as an adaptive filtering problem, exploiting the NLMS algorithm to update a_k and b_k in the following way:

$$\boldsymbol{w}(n+1)=\boldsymbol{w}(n)+\mu e_{\text{est}}(n)\frac{\boldsymbol{x}(n)}{\boldsymbol{x}^H(n)\boldsymbol{x}(n)+\psi}, \quad (7)$$

where $\boldsymbol{w}(n)=[a_1(n)\cdots a_M(n),\ b_{-Q}(n)\cdots b_Q(n)]^T$ and $\boldsymbol{x}(n)=[x(n-D)\cdots x(n-M-D+1),\ x(n-T+Q)\cdots x(n-T-Q)]^T$ while μ and ψ are step size and regularization parameters, respectively.

4. SIMULATION RESULTS

In this section we firstly demonstrate multi-step prediction performance of joint STP and LTP, and LTP alone, finding the trade-off between computational complexity (filter order) and prediction performance. Then simulations for the proposed fixed-filter prediction-based ANC system, with comparisons to conventional adaptive ANC, are presented.

For the following simulations, 12 male and female speech signals with an average length of 3 seconds, sampled at 8 kHz from the ITU-T database [21] were used. The estimation of T for the LTP was performed based on [22] with T being estimated on a sample-by-sample basis. Other simulation conditions are: $P(z)$, $S(z)$ and secondary path model $\check{S}(z)$ (for the adaptive FXNLMS ANC) with a length of 150 were simulated as band-pass minimum phase filters that satisfy the causality constraint. The pass-band frequency range for $P(z)$ is $(100, 2000)$, and $(80, 3500)$ for $S(z)$ and $\check{S}(z)$, respectively. Regularization parameter for the adaptive ANC and LTP, $\psi=10^{-3}$; step size μ equal to 0.045 and 0.07, respectively, and it was found based on empirical simulations defining the best possible performance of both systems.

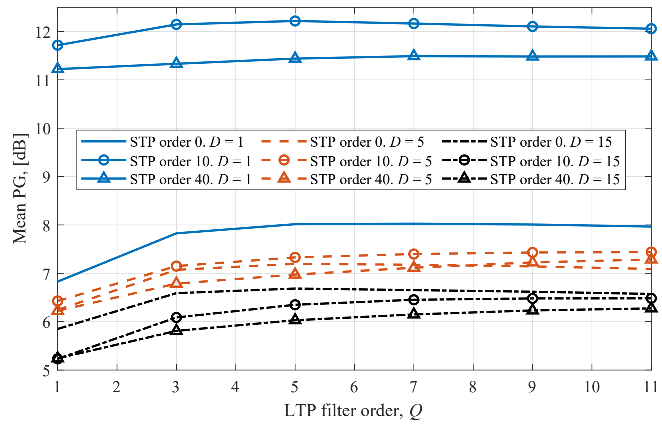

Fig. 3. Mean PG of voiced parts for multi-step prediction D by STP and LTP as a function of LTP order. Averaged over 12 speech signals.

4.1. Speech prediction

To determine the optimal order of the combined STP and LTP for voiced speech, the classical one-step prediction scheme along with multi-step linear prediction case were investigated, as reported next. As a performance metric, the prediction gain (PG) [18] measured in dB, was used and calculated over a sliding window, i.e.,

$$PG(n)=10\log_{10}\left(\frac{\sum_{n=-I}^{I} x(n)^2}{\sum_{n=-I}^{I}(x(n)-\hat{x}(n))^2}\right), \quad (8)$$

where $2I+1$ is the length of the sliding window. A better predictor is capable of generating lower prediction error, leading to a higher prediction gain. The PG was calculated with a sliding window of 25 ms, then averaged over speech samples with high voicing probability (>0.9). For this purpose the voicing-unvoicing detection based on the fast pitch tracking algorithm using the harmonic model was used [23].

The PG averaged over 12 speech signals for joint STP and LTP, and LTP only (corresponding to an STP order of 0) is presented in Fig. 3. Prediction is performed for 1, 5 and 15 samples (steps) ahead. As can be seen, for the classical LP scheme with one-step prediction, a combination of STP of order 10 with LTP gives an increase in PG of about 4 dB compared to LTP only. Further increase in the order of the STP to 40 leads to a slight decrease of PG. This experiment is in line with previous experiments reported in [18]. However, when it comes to multi-step prediction with 5 and 15 samples ahead, adding STP to LTP generally decreases mean PG except for a negligible improvement for $D=5$ and STP order 10. Thereby, for the proposed ANC system, and further simulations, STP of order 0 and LTP of order 5 is used, which seems to be optimal with respect to the performance and computational complexity. The Prediction performance of the voiced speech in our study depends on several factors, such as the accuracy of T estimation, with its possible fractional value, rounded here, degree of the speech non-stationarity, and the accuracy of the used voiced speech detection.

4.2. Speech attenuation

The effectiveness of the proposed fixed-filter FF ANC system in Fig. 2 with integrated LTP of order 5 is demonstrated in this subsection. The proposed system is also compared with a conventional

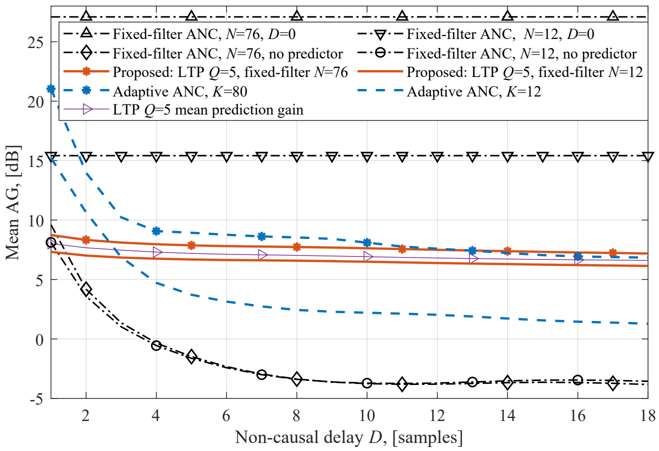

Fig. 4. Mean AG of voiced parts as a function of the non-causal delay D. Averaged over 12 speech signals.

FF FXNLMS ANC system [2] in terms of attenuation gain (AG) in dB, which is calculated in a similar way as PG in (8).

$$AG(n) = 10\log_{10}\left(\frac{\sum_{n=-I}^{I} d(n)^2}{\sum_{n=-I}^{I}(d(n)-\hat{y}'(n))^2}\right). \tag{9}$$

Voiced parts of speech were taken into account here as well for calculating mean AG.

Fig. 4 compares the speech attenuation performance of the proposed system with an adaptive ANC system for different amounts of non-causal delay, D. For comparison, two orders of the fixed-filter and adaptive filter were used. The choice for the fixed-filter order $N = 76$ and AF order $K = 80$ taps is based on the empirical simulations defining the best possible performance of the systems for all non-causal delays. Further increase of the fixed-filter order up to the same length as $P(z)$ does not bring any performance improvement for both, i.e., the proposed system and fixed-filter ANC without predictor. The order of 12 is presented for comparison and addresses the question of what can be achieved with a much shorter filter.

The performance of the causal fixed-filter ANC system (without predictor) shows quite high speech mean AG of about 15 and 27 dB for filter length $N = 12$ and $N = 76$, respectively. This might be interpreted as an upper bound of the fixed-filter ANC system attenuation performance, and in the case of exploiting an ideal multi-step predictor, attenuation performance will not be affected by the non-causal delay D. However, even a delay of one sample leads to a drop in performance of about 8 and 17 dB, respectively, resulting in almost indistinguishable performances for the two filter lengths. Further delay increase makes the fixed-filter ANC system without predictor ineffective to attenuate speech, leading even to speech amplification, which is indicated by the negative mean AG.

From Fig. 4, we observe that introducing the LTP in the fixed-filter ANC system, provides significant improvement with a fairly uniform speech attenuation performance over a wide range of non-causal delays. The difference between mean AG for the fixed-filter order of 12 and 76 is about 1.2 dB only, and taking into account the mean PG of the LTP, one can conclude, that LTP performance is directly related to, or defines speech attenuation performance of the proposed system. This means that a predictor with higher performance will increase mean AG of the proposed system. Comparing speech attenuation performance to the adaptive ANC system, it can

Table 1. Computational complexity in algorithms. K, $L\hat{s}$, Q and N are the filter orders of the AF, $\check{S}(z)$, LTP and fixed filter, respectively.

Algorithm		Multiplication	Addition	Division
Adaptive ANC, FXNLMS		$3K+L\hat{s}+1$	$3K+L\hat{s}-3$	$L\hat{s}$
Proposed system:	1. LTP, NLMS	$8Q+5$	$8Q+3$	$2Q+1$
	2. Fixed-filter	$2N+1$	$2N$	-

be seen that for delays up to 3 samples, which corresponds to 0.375 ms, the adaptive ANC system ($K = 80$) has a noticeable advantage in the mean AG, but with further increase of non-causal delay, only a higher order of the AF gives a comparable performance with the proposed ANC system. At high delays, the proposed system slightly outperforms the adaptive ANC system. Comparing the systems in terms of the adaptation process reveals some interesting things. In the LTP the goal is to minimize the estimation error in (9) and then to perform prediction (5). While in the adaptive ANC system, residual ANC error (error at the desired point) is minimized, which also includes prediction error in the sense that the AF output signal should also be predicted to align with the primary disturbance in case of non-causal delay.

4.3. Computational complexity

The computational complexity of the algorithms and corresponding systems in terms of mathematical operations is summarized in Table 1. We remark that the computational cost related to estimating T in the LTP is not taken into account here, as it can be estimated fast in a number of ways, for example using [22,23] every 15–40 ms combined with interpolation of the so-obtained estimates. For the performance shown in Fig. 4 the adaptive ANC system with an order of 80 would require $\{391, 387, 80\}$ number of multiplications, additions and division respectively, while the proposed ANC system with LTP of order 5, and fixed-filter order of 12 and 76 taps requires $\{46, 43, 5\}$ and $\{174, 171, 5\}$ number of corresponding operations, respectively. Thus, the proposed ANC system outperforms conventional adaptive ANC systems in terms of computational cost, providing comparable performance in speech attenuation for non-causal delays more than 3 samples, and, especially, at higher delays.

5. CONCLUSION

In this paper, we proposed a feedforward ANC system based on a fixed-filter with integrated long-term linear prediction to solve the issue of non-causal delay, when the noise to be cancelled is speech. The performance of multi-step prediction for voiced speech obtained with joint short- and long-term prediction was investigated, where the results showed that the use of LTP only without STP is the best choice. The effectiveness of the proposed ANC system was confirmed through simulations using 12 speech signals from the ITU-T database. Simulations show that the proposed ANC system reaches comparable or better mean attenuation gain of voiced speech, which is on average 8 dB, than the conventional adaptive feedforward FXNLMS ANC system at a wide range of non-causal delays, i.e., 4 to 18 samples (0.5 to 2.25 ms) at a sampling frequency of 8 kHz, while reducing the computational complexity. Future work should focus on a more effective prediction scheme leading to higher prediction performance, since simulations show that the performance of the proposed system is linked to the prediction performance.

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] 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.

[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, 2012.

[4] S. M. Kuo, S. Mitra, and W.-S. Gan, "Active noise control system for headphone applications," *IEEE Transactions on Control Systems Technology*, vol. 14, no. 2, pp. 331–335, 2006.

[5] T. Schumacher, H. Krüger, M. Jeub, P. Vary, and C. Beaugeant, "Active noise control in headsets: A new approach for broadband feedback anc," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 2011, pp. 417–420.

[6] S. Priese, C. Bruhnken, D. Voss, J. Peissig, and E. Reithmeier, "Adaptive feedforward control for active noise cancellation in-ear headphones," *J. Acoust. Soc. Am.. Proceedings of Meetings on Acoustics*, vol. 18, no. 1, 2012.

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

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

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] M. Nishimura, T. Tanaka, K. Shiratori, K. Sakurama, and S. Nishida, "Development of a voice shutter (phase 1: A closed type with feed forward control)," in *Proc. INTER-NOISE*, 2014, vol. 249, pp. 1291–1299.

[15] K. Kondo and K. Nakagawa, "Speech emission control using active cancellation," *Speech Communication*, vol. 49, pp. 687–696, 09 2007.

[16] D. Suzuki and K. Kondo, "Application of anc for singing voice attenuation," in *2014 IEEE Global Conference on Consumer Electronics*, 2014, pp. 63–64.

[17] S.-W. Jeon, D. H. Youn, Y.-C. Park, and G.-W. Lee, "Active control of excessive sound emission on a mobile device," *J. Acoust. Soc. Am.*, vol. 137, no. 4, pp. 327–333, 2015.

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

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

[20] S. V. Vaseghi, *Advanced digital signal processing and noise reduction.*, Wiley, Chichester, 2. ed. edition, 2000.

[21] "Recommendation ITU-T P.501: Test signals for use in telephony and other speech-based applications," .

[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] 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/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 11, pp. 1737–1751, 2019.