ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 979-8-3503-4485-1/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICASSP48485.2024.10447674

IMPROVING SPEECH ATTENUATION IN HEADPHONES USING HARMONIC MODEL DECOMPOSITION AND MULTIPLE-FREQUENCY ANC

Yurii Iotov[∗1,2], *Sidsel Marie Nørholm*[2], *Peter John McCutcheon*[2], *Mads Græsbøll Christensen*[1]

[1] Audio Analysis Lab, Dept. of Electronic Systems, Aalborg University, Denmark, {yio, mgc}@es.aau.dk
[2] GN Audio A/S, Ballerup, Denmark, {yiotov, snoerholm, pjmccutcheon}@jabra.com

ABSTRACT

In environments such as open offices, call centres, etc., speech is often the main disturbing source of ambient noise, reducing concentration and productivity. 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. To deal with this, we propose a new feedforward ANC system for headphone applications, HMD-ANC, which improves voiced speech attenuation. Notably, in HMD-ANC, each speech harmonic and its quadrature version, obtained by the harmonic model decomposition, are predicted by a two-weight adaptive filter to overcome the delay. The results show that HMD-ANC outperforms conventional adaptive feedforward ANC for delays starting from 6 samples (0.125 ms) at a sampling frequency of 48 kHz. Moreover, HMD-ANC extends the attenuation bandwidth, e.g., up to 2.5 kHz, while the conventional ANC is limited to 1 kHz.

Index Terms— ANC, causality, speech prediction, harmonics.

1. INTRODUCTION

ANC has been successfully applied in the past to deal with different types of noise in many applications [1–5]. Nowadays, ANC is becoming a key feature of modern headphones helping to reduce loud and annoying noise from transportation, aircraft or train noise when travelling or commuting. However, in environments such as open offices, call centres, and co-working places, speech is often the main disturbing source of ambient noise, reducing concentration and productivity. Hence, it is increasingly important that ANC headphones also attenuate speech as effectively as other types of noise. However, speech attenuation by ANC headphones can be quite limited due to the complex nature of speech and constraints in the ANC system.

The main principle behind ANC is the destructive interference of sound waves. Using as an example the feedforward (FF) ANC headphones in Fig. 1, to cancel unwanted noise at the eardrum, an anti-noise signal with the same amplitude but the opposite phase is generated by a headphone loudspeaker. Many factors affect ANC performance in headphones, including the causality constraint [2, 3, 6–14]. For FF ANC headphones, in Fig. 1, the system is causal when the sum of the electric delay ED_S and acoustic propagation delay AD_S in the secondary path $S(z)$ is less than or equal to the acoustic propagation delay of noise, AD_P, through the primary path $P(z)$. However, the causality constraint is often violated due to an additional delay in $S(z)$ relative to $P(z)$, $D = ED_S + AD_S - AD_P$. This might

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

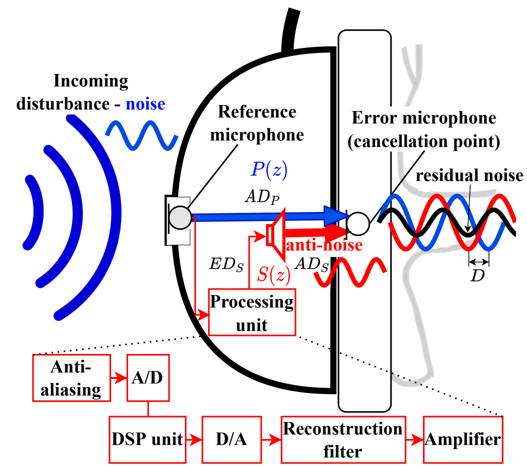

Fig. 1. Simplified modelling block diagram of FF ANC headphones.

be 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, which is a very device-specific parameter and depends on the ANC processing unit and its algorithmic design [11–13], can also be a cause of the delay D. The direction of the incoming noise [8] and improper headphone fit on the ear [6] can also affect D.

When the causality constraint is violated, it creates the need for prediction to compensate for the delay D [7]. Conventional adaptive ANC algorithms, i.e., Filtered-X Normalized least mean square (FXNLMS), act as a linear predictor to find a causal ANC filter [10, 15]. Hence, the performance of such an ANC system depends on the predictability of the noise and its bandwidth [8, 10]. When the noise to be attenuated is speech, the problem of compensating for the delay D, i.e., predicting future speech samples, becomes quite complex. This is because speech tends to be non-stationary and has a complex structure broadly divided into quasi-periodic voiced speech and unvoiced speech, which is stochastic in nature and almost unpredictable [16]. With a structure of a fundamental frequency and a set of overtones given by integer multiples of the fundamental, voiced speech is the main constituent of speech [17]. Moreover, the frequency range of voiced speech is in the range from 80 Hz to 4-5 kHz, but also reaches beyond [16, 18]. In contrast, the energy of unvoiced speech usually concentrates at higher frequencies, e.g., above 3-4 kHz [18, 19], which can be efficiently reduced by the passive attenuation of the headphones. Therefore, in the context of ANC, the primary focus of attenuation is voiced speech.

When conventional adaptive ANC is employed for speech attenuation, the FXNLMS algorithm, acting as a linear predictor (LP) to compensate for D, has a suboptimal filter structure for voiced speech prediction, which leads to a limited mean attenuation level and bandwidth [20]. This is because only the filter coefficients corresponding to the short- and long-term correlations of voiced speech contribute positively to the prediction performance [16]. Therefore, the joint short-long term predictor (SLTPj) is a better LP scheme for prediction in FF ANC, outperforming FXNLMS ANC for a wide range of the delay D by up to 2.5 dB in voiced speech attenuation [20]. However, FXNLMS ANC and SLTPj ANC systems have quite poor performance at frequencies above 1-1.5 kHz [20] where the need for prediction is greater as a constant time delay constitutes a larger phase difference at a higher frequency. In addition, the human ear is most sensitive to sounds around 2–4 kHz [18]. Hence, it is critical to find a solution which will improve voiced speech attenuation and increase the prediction frequency range towards higher frequencies when compensating for the delay D.

Inspired by the idea of multiple-frequency ANC [2, 21, 22], which has been efficiently applied to reduce multi-tone disturbances, e.g., engine-generated noise [2, 23], and the structure of quasi-periodic voiced speech, this paper investigates its application to voiced speech when compensating for the delay D. We propose a new parallel-form multiple-frequency FF ANC system, HMD-ANC, where each speech harmonic and its quadrature version (with a 90°phase shift) are used as reference inputs for the two-weight adaptive FXNLMS filter. Specifically, we propose to decompose voiced speech into harmonics by the harmonic model (HM), which has been successfully applied to a broad range of speech processing problems [24, 25]. As was demonstrated in [26], it is possible to predict a sample composed of the sum of L sinusoids as a linear combination of its previous $2L$ samples. Hence, a quasi-periodic speech harmonic, exhibiting similarities to a sinusoid, should also be predicted by the two-weight adaptive filter; however, not as good as a periodic and stationary tone.

This work represents a new idea and proof of concept for improving voiced speech attenuation while compensating for the delay D. Therefore, other challenges inherent in headphone ANC, such as changes in $P(z)$ and $S(z)$, reference-error microphone coherence, the effect of reverberation and other background noise, are not considered and exceed the scope of the paper. This allows for breaking down the complexity of an ANC system, focusing on the primary scope and establishing the upper-bound performance of the proposed method.

The paper is organised as follows. First, HMD-ANC, based on harmonic model speech decomposition and the multiple-frequency ANC, is described in Section 2. Simulation results of voiced speech attenuation are presented in Section 3. Section 4 concludes the paper.

2. PROPOSED DECOMPOSITION-BASED ANC SYSTEM

The proposed FF HMD-ANC system, shown in Fig. 2, is designed to improve voiced speech attenuation. It is intended to work alongside with, e.g., a conventional ANC system to attenuate other types of noise, including, unvoiced speech. However, this paper focuses only on HMD-ANC. The diagram in Fig. 2 is for a single headphone, which uses one reference and one error microphone to measure the incoming noise $x(n)$, i.e., speech, and the residual error $e(n)$. The goal is to match the internal disturbance $d(n)=x(n)*p(n)$, where $*$ denotes linear convolution, $p(n)$ is the time domain version of $P(z)$, as accurately as possible in amplitude and inverted phase. The proposed system can be conceptually divided into speech decomposition and ANC parts.

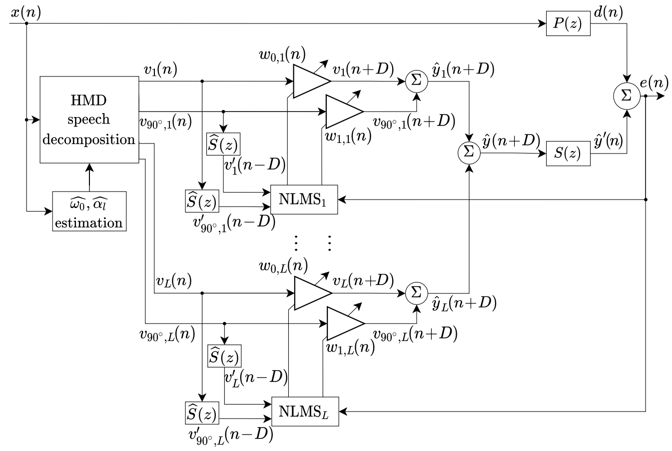

Fig. 2. Block diagram of the proposed FF HMD-ANC system.

2.1. Speech decomposition

A single-speaker speech signal is modelled by $x(n) = v(n) + u(n)$, where $v(n)$ is the harmonic part of the signal, i.e., voiced speech, $u(n)$ includes the stochastic part, i.e., unvoiced speech, background noise, and model error. The harmonic part can be represented as $v(n) = \sum_{l=1}^{L} v_l(n)$, where L is the total number of harmonics and $v_l(n)$ is the l'th harmonic. The harmonic $v_l(n)$ with its quadrature version $v_{90^\circ,l}(n)$ are given by the real harmonic model as [17]:

$$v_l(n) = A_l \cos(l\omega_0 n + \phi_l), \tag{1}$$

$$v_{90^\circ,l}(n) = A_l \sin(l\omega_0 n + \phi_l), \tag{2}$$

where A_l is a real, non-zero amplitude, $\omega_0 = 2\pi f_0/f_\mathrm{s}$, with f_0 the fundamental frequency and f_s the sampling frequency, is the normalized fundamental frequency, and $\phi_l \in [0, 2\pi)$ is the initial phase of the l'th harmonic [17]. In the harmonic model, the instantaneous frequency ω_0 is assumed stationary within a segment of samples. The estimates of the harmonic model parameters denoted with $(\widehat{\cdot})$, $\widehat{\omega}_0$ and $\widehat{\boldsymbol{\alpha}} = [\widehat{A_1}, \dots \widehat{A_L}, \widehat{\phi_1} \dots \widehat{\phi_L}]$, are found by minimizing the mean square error given by $\mathbb{E}\{|e_M(n)|^2\} = \mathbb{E}\{|x(n) - \widehat{v}(n)|^2\}$. This can be done in a number of ways, e.g., the non-linear least-squares (NLS) method has proven to be one of the most accurate and robust [27]. A Kalman filter, similar to [28,29], or the recursive least squares (RLS) or LMS algorithms, can also be used to update the estimates. Considering non-stationarity and quasi-periodicity of speech harmonics, $\widehat{\omega}_0$ and $\widehat{\boldsymbol{\alpha}}$ should be updated regularly.

2.2. ANC

For each speech harmonic, the signals $v_l(n)$ and $v_{90^\circ l}(n)$, used as reference inputs for the 2-tap l'th FXNLMS adaptive filter, are separately weighted and then summed to produce the cancelling signal

$$\hat{y}_l(n + D) = v_l(n) w_{0,l}(n) + v_{90^\circ l}(n) w_{1,l}(n), \tag{3}$$

where the time index $(n + D)$ denotes prediction by the adaptive filter with the weights w updated by the FXNLMS algorithm as:

$$w_{0,l}(n + 1) = w_{0,l}(n) + \mu e(n) \frac{v_l'(n-D)}{\|v_l'(n-D)\|^2 + \psi}, \tag{4}$$

$$w_{1,l}(n + 1) = w_{1,l}(n) + \mu e(n) \frac{v_{90^\circ l}'(n-D)}{\|v_{90^\circ l}'(n-D)\|^2 + \psi}, \tag{5}$$

where $v'_l(n-D)$ and $v'_{90^\circ l}(n-D)$ are $v_l(n)$ and $v_{90^\circ l}(n)$, respectively, filtered through the secondary path estimate $\widehat{S}(z)$, which contains the delay D, μ is the step-size, and ψ is a small positive number. In HMD-ANC, the full-band error $e(n)$ common to all L adaptive filters is used. As demonstrated in [30], the use of full-band error performs equally well as the narrow-band error in cancelling sinusoids without a negative effect on the adaptation process of the l'th adaptive filter.

The steady-state transfer function between $d(n)$ and $e(n)$, $H(z)$, has the property of a notch filter [2]. For example, when $x(n)$ is defined as $v_l(n)$ in (1) for $l=1$, and for simplicity, $\widehat{S}(z)=S(z)$ is assumed to be a pure delay, $H(z)$ is given as [31]

$$H(z)=\frac{z^2-2z\cos(\omega_0)+1}{z^2-2z\cos(\omega_0)+1+\mu A^2 S(z)[z\cos(\omega_0-\Phi_s)-\cos(\Phi_s)]}, \tag{6}$$

where $\Phi_s=-D\omega_0$ represents a phase shift, i.e., the delay D. The notch is located at ω_0, and the bandwidth of the notch filter is a function of D and μ, with a larger μ leading to faster tracking and a wider notch [31]. The idea of generating $v_{90^\circ l}(n)$ in (2) for the reference input in (3) allows for higher precision of the notch filter, i.e., making it sharper at ω_0 while improving the convergence rate [2,21]. For non-stationary speech, faster convergence is advantageous. Considering the complex and time-varying nature of speech harmonics, we can assume that $H(z)$ for the l'th FXNLMS adaptive filter in HMD-ANC, similar to the analysis in [21], will tend towards (6). For L speech harmonics, it will represent a transfer function of a filter with multiple notches, e.g., as in [21].

Based on the above, the cancelling signal for L harmonics is given as $\hat{y}(n+D)=\sum_{l=1}^{L}\hat{y}_l(n+D)$, where the predicted harmonics compensate for the delay D in $S(z)$, resulting in $d(n)$ and $\hat{y}'(n)$ being aligned in time at the error microphone, so the residual error $e(n)$ is minimized. Hence, the higher the accuracy of the predicted speech harmonics $\hat{y}(n+D)$, the lower the residual error and the better the voiced speech attenuation, which was also shown in [20].

3. SIMULATION RESULTS

3.1. Simulation conditions

For the following simulations, $P(z)$ and $S(z)$ shown in Fig. 3 were measured on a Jabra headphone with a head and torso simulator. The measurements were done in an anechoic chamber with a point source representing the ambient noise to ensure high coherence between the reference and error microphones, with $S(z)$ excluding the ANC processing unit. Considering the factors which can violate the causality constraint discussed in Section 1, e.g., the latency of the ANC processing unit, which can be from tens to thousands of microseconds [12], 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, we used $P_{\mathrm{m}}(z)$ for the primary path and $S_{\mathrm{m}}(z)z^{-D}$ for the secondary path, where the subscript m denotes the minimum-phase parts of $P(z)$ and $S(z)$ shown in Fig. 3, calculated with the real cepstrum method [32].

As an ambient noise input to the system, i.e., $x(n)$ in Fig. 2, we used 10 female and 10 male utterances from different speakers, each with an average duration of 2.5 s, from the TSP speech database recorded in an anechoic room with $f_{\mathrm{s}}=48$ kHz [33]. The simulation parameters listed below were found empirically. They maximize the performance of systems under the given conditions, which allows us to make a fair comparison. When estimating $\widehat{\omega}_0$ and $\widehat{\boldsymbol{\alpha}}$, there is a trade-off between segment length and accuracy of the estimates [27].

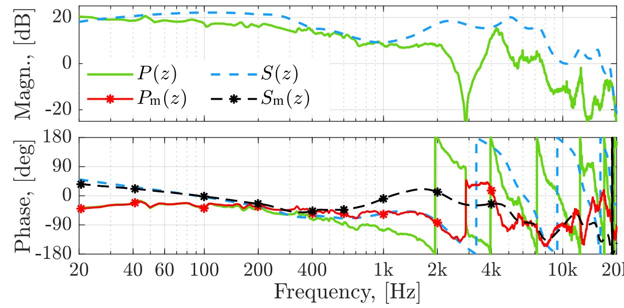

Fig. 3. Measured $P(z)$, $S(z)$ and their minimum-phase parts.

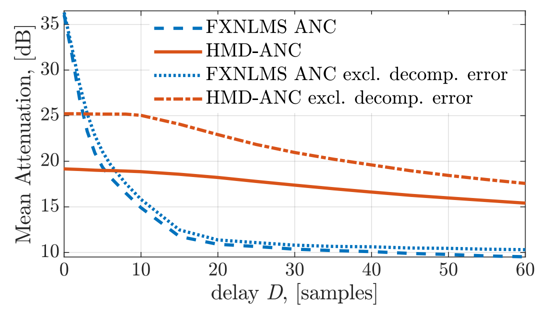

Fig. 4. Mean attenuation of voiced speech as a function of D.

The parameter $\widehat{\omega}_0$ was estimated every 6 samples using the NLS estimator from [27] with the segment length of 20 ms and a model order of $L=20$. The amplitudes and initial phases $\widehat{\boldsymbol{\alpha}}$ were estimated every sample with the least squares method [34] over a window of 20 ms and including all harmonics up to $f_{\mathrm{s}}/2$. However, for HMD-ANC, L was set to include a maximum of 40 speech harmonics, which results in 40 parallel 2-tap FXNLMS filters. Depending on f_0, this will correspond to the upper-frequency range of 4.5-12 kHz and will usually cover all the prominent speech harmonics [18]. In general, L can be chosen depending on, e.g., the passive attenuation of the headphones and the desired ANC effective frequency range. The impulse response length for $P(z)$, $S(z)$ and $\widehat{S}(z)$ is 600. For HMD-ANC, μ is 0.005. The conventional adaptive FF ANC system [2], FXNLMS ANC, with the order of 600 and $\mu=0.028$, serves as the baseline for comparison. For all systems, $\psi=10^{-3}$. For performance evaluation, the attenuation metric A, in (7), was calculated on voiced speech samples, detected by the method in [35], over a sliding window of 12 ms. The higher the A metric, the better the ANC performance.

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

3.2. Results

The average voiced speech attenuation performance when compensating for the delay D is shown in Fig. 4. Comparing the performance of conventional FXNLMS ANC to HMD-ANC, it can be seen that when the system is causal, i.e., $D=0$, FXNLMS ANC shows quite high mean attenuation of about 35 dB, while HMD-ANC reaches around 20 dB of attenuation. Although an average attenuation of 35 dB may not be achievable in practice due to other factors affecting ANC performance, the performance difference can be explained by the ability of FXNLMS ANC with a long filter to

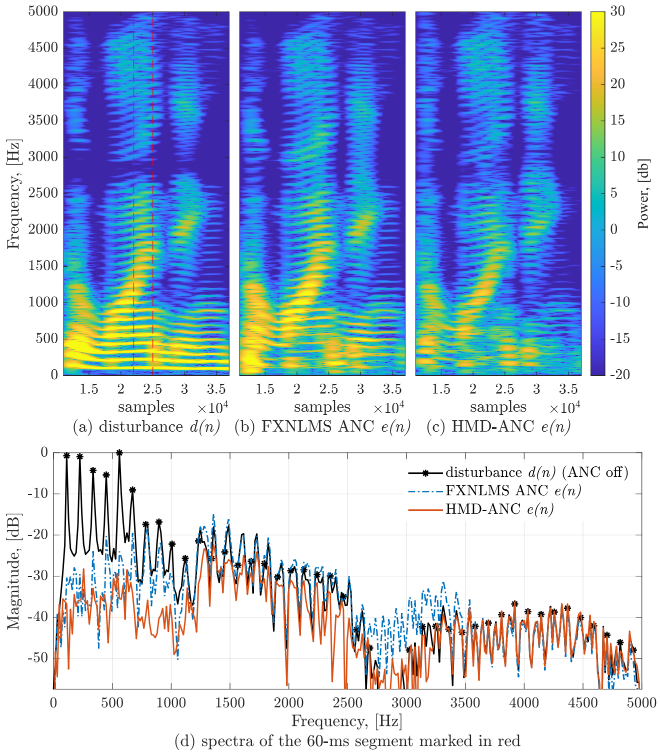

Fig. 5. Spectrogram of (a) $d(n)$—male voiced speech segment and $e(n)$ for (b) FXNLMS ANC and (c) HMD-ANC compensating for the delay D of 20 samples; (d) shows spectra of a smaller segment.

better model the phase and magnitude of the causal FF ANC filter. However, FXNLMS ANC attenuation performance degrades significantly as the delay increases, e.g., it drops by a factor of two at $D=5$ and by a factor of three at $D=15$ samples, where it also approaches its lowest performance. In contrast, the proposed HMD-ANC shows relatively stable performance in the wide range of D, outperforming FXNLMS ANC for $D \geq 6$ samples. Specifically, HMD-ANC improves voiced speech attenuation by at least 3.0 dB and up to 7.3 dB over FXNLMS ANC in the range of $9 \leq D \leq 60$ samples.

To evaluate the performance of the multiple-frequency ANC approach applied for voiced speech attenuation, Fig. 4 also shows plots for FXNLMS ANC and HMD-ANC excluding the harmonic model decomposition error, i.e., $d(n) = v(n) * p(n)$. This allows us to define the approximate upper-bound performance of the proposed system and what can be achieved with better speech decomposition. In such a case, HMD-ANC outperforms FXNLMS ANC for $D \geq 3$, by 4.0 dB at $D=5$, and up to 11.5 dB at $D=20$ samples.

For a more detailed performance evaluation, Fig. 5 shows spectrograms for a segment of male voiced speech disturbance (a), and the corresponding residual error for (b) FXNLMS ANC and (c) HMD-ANC for $D=20$ samples where the maximum performance difference is achieved. The shown example of a speech segment represents a complex case since it illustrates the non-stationarity of speech and contains many pronounced harmonics with higher harmonics showing relatively high amplitude modulation, i.e., change in amplitude over time. Compared to FXNLMS ANC, HMD-ANC indicates better attenuation up to around 1.5 kHz and is still effective up to 2.5 kHz, as is seen in Fig. 5(d). In contrast, FXNLMS ANC has a larger error and is only effective up to 1 kHz with a quite poor performance above 500 Hz. In the 2.5-3.5 kHz range, FXNLMS ANC significantly amplifies speech rather than attenuating it.

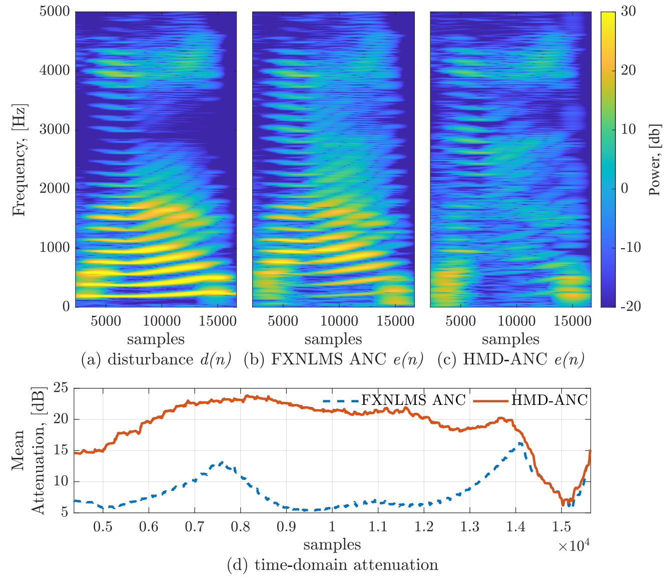

Fig. 6. Spectrogram of (a) $d(n)$—female voiced speech segment and $e(n)$ for (b) FXNLMS ANC and (c) HMD-ANC compensating for $D=20$ samples; (d) shows corresponding time-domain attenuation.

Another example of the attenuation performance for a female voiced speech segment is shown in Fig. 6, where the non-stationarity of speech is more pronounced, i.e., the frequencies of the harmonics are changing with a noticeable amount of frequency modulations. In this case, HMD-ANC shows a quite prominent improvement in attenuation compared to FXNLMS ANC up to 2.5 kHz, which can be seen by comparing Fig. 6(b) and Fig. 6(c). This results in up to 15 dB higher attenuation of that speech segment, as shown in Fig. 6(d).

Although subjective tests are planned for the future, the results from an informal listening test[1] are mainly in line with the results discussed in this section and generally point to a noticeable improvement in attenuation by the proposed HMD-ANC at the aforementioned delays where it outperforms FXNLMS ANC system.

4. CONCLUSION

We proposed a new FF ANC system for headphone applications, HMD-ANC, which improves voiced speech attenuation when compensating for the delay D violating the causality constraint. Specifically, in HMD-ANC, each speech harmonic and its quadrature version, obtained by the harmonic model decomposition, are predicted by a two-tap FXNLMS filter updated with the full-band error. Simulations show that HMD-ANC outperforms conventional adaptive FF FXNLMS ANC for $D \geq 6$ samples (≥ 0.125 ms) at an f_s of 48 kHz with an improvement of 3.0-7.3 dB in the range of $9 \leq D \leq 60$ samples (0.1875 to 1.25 ms). Moreover, HMD-ANC extends the attenuation bandwidth, e.g., up to 2.5 kHz at D of 20 samples, where FXNLMS ANC is only effective up to 1 kHz with a quite poor attenuation above 500 Hz. To also attenuate other types of noise, HMD-ANC can work alongside with, e.g., FXNLMS ANC, which can also be applied to speech when the delay changes to $D < 6$ samples.

[1]some audio examples are available at `https://www.dropbox.com/sh/q17sjv5q0s1rzc6/AAA0mwtB8fCRk6tTMJlsrfp0a`

5. 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," *Proc. 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 Trans. Signal Inf. Process.*, vol. 1, pp. 1–21, 2012.

[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 I: Linear systems," *Signal Process.*, vol. 183, pp. 108039, 2021.

[5] 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 Process.*, vol. 181, pp. 107929, 2021.

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

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

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

[9] 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," *J. Sound Vib.*, vol. 141, no. 3, pp. 409–424, 1990.

[10] 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.

[11] 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.

[12] S. Liebich, J. Fabry, P. Jax, and P. Vary, "Signal processing challenges for active noise cancellation headphones," in *Proc. Speech Commun. 13th ITG-Symp.*, 2018, pp. 1–5.

[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," *Appl. Acoust.*, vol. 179, 2021.

[14] X. Shen, D. Shi, and W.-S. Gan, "A hybrid approach to combine wireless and earcup microphones for anc headphones with error separation module," in *Proc. IEEE ICASSP*, 2022, pp. 8702–8706.

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

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

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

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

[19] B. B. Monson, A. J. Lotto, and B. H. Story, "Analysis of high-frequency energy in long-term average spectra of singing, speech, and voiceless fricatives," *J. Acoust. Soc. Am.*, vol. 132, no. 3, pp. 1754–1764, 09 2012.

[20] Y. Iotov, S. M. Nørholm, V. Belyi, and M. G. Christensen, "Non-stationary prediction for addressing the non-causality problem in fixed-filter ANC headphones for speech reduction," in *Proc. IEEE EUSIPCO*, 2023, pp. 101–105.

[21] J. Glover, "Adaptive noise canceling applied to sinusoidal interferences," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 25, no. 6, pp. 484–491, 1977.

[22] W. Zhu, L. Luo, M. G. Christensen, and J. Sun, "A new feedforward hybrid active control system for attenuating multi-frequency noise with bursty interference," *Mech. Syst. Signal Process.*, vol. 144, pp. 106859, 2020.

[23] T. Wang, W.-S. Gan, and S.M. Kuo, "New feedback active noise control system with improved performance," in *Proc. IEEE ICASSP*, 2014, pp. 6662–6666.

[24] 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.

[25] 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.

[26] 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.

[27] 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 Process.*, vol. 135, pp. 188–197, 2017.

[28] M. G. Christensen, "A method for low-delay pitch tracking and smoothing," in *Proc. IEEE ICASSP*, 2012, pp. 345–348.

[29] 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 WASPAA*, 2017, pp. 314–318.

[30] J. He, W.-S. Gan, and Y.-K. Chong, "Study on the use of error term in parallel-form narrowband feedback active noise control systems," in *Proc. Asia-Pacific. Signal Inf. Process. Assoc. (APSIPA) Annu. Summit Conf.*, 2014, pp. 1–6.

[31] Z. Limin, T. Jiancheng, and Q. Xiaojun, "Active control of transformer noise with an internally synthesized reference signal," *J. Sound Vib.*, vol. 331, no. 15, pp. 3466–3475, 2012.

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

[33] P. Kabal, "TSP Speech Database," Tech. Rep., Department of Electrical and Computer Engineering, McGill University, 2018.

[34] M. G. Christensen and A. Jakobsson, *Multi-Pitch Estimation*, vol. 5 of *Synthesis Lectures on Speech and Audio Processing*, Morgan & Claypool Publishers, San Rafael, CA, 2009.

[35] 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.