INTERSPEECH 2016
September 8–12, 2016, San Francisco, USA

A Spectral Modulation Sensitivity Weighted Pre-emphasis Filter for Active Noise Control System

Kah-Meng Cheong [1], *Yuh-Yuan Wang* [2] *and Tai-Shih Chi* [1]

[1] Department of Electrical and Computer Engineering
National Chiao Tung Universty, Hsinchu, Taiwan 300, R.O.C.
[2] TANGENT Microelectromechanics Corp., Taiwan

Abstract

Psychoacoustic active noise control (ANC) systems by considering human hearing thresholds in different frequency bands were developed in the past. Besides the frequency sensitivity, human hearing also shows different sensitivity to spectral and temporal modulations of the sound. In this paper, we propose a new psychoacoustic active noise control system by further considering the spectral modulation sensitivity of human hearing. In addition to the sound pressure level (SPL), the loudness level is also objectively assessed to evaluate the noise reduction performance of the proposed ANC system. Simulation results demonstrate the proposed system outperforms two compared systems under test conditions of narrowband and broadband noise in terms of the loudness level. The proposed algorithm has been validated on TI C6713 DSP platform for real-time process.

Index Terms: Psychoacoustic active noise control, spectral modulation, hearing threshold, loudness

1. Introduction

The active noise control (ANC) is a well-developed technique for reducing noise by generating and combining an appropriate anti-noise signal [1]. The filtered-x least-mean-square (FXLMS) algorithm is the most commonly used algorithm due to its simplicity. Many algorithms have been proposed over the past decades to improve the convergence speed and the overall performance of ANC systems from the engineering point of view [2]. Since outputs of ANC systems are for human ears, it is intuitive to consider hearing properties in developing ANC systems. The filtered-error least-mean-square (FELMS) algorithm, which incorporates different noise weights to shape the spectrum of residual error, was proposed in [3] and motivated the development of psychoacoustic ANC systems. As a result, psychoacoustic ANC systems, which consider the hearing thresholds at different acoustic frequencies, were proposed in the FELMS framework over the past few years [4][5].

In addition to different sensitivities to different acoustic frequencies, the human brain also analyzes sounds in terms of their joint spectro-temporal modulations [6]. This concept of emphasizing different spectro-temporal modulations has been successfully applied in many speech-related applications, such as noise reduction [7], voice activity detection [8], and speech intelligibility assessment [9]. The detection thresholds of joint spectro-temporal modulations have been measured in psychoacoustic experiments and demonstrated separable in terms of detection thresholds of pure spectral and pure temporal modulations [10]. In other words, human hearing possesses different sensitivities to different spectral and temporal modulation frequencies of the sound.

For a real-time application such as ANC, considering temporal modulation across a period of time is not practical. Therefore, in this paper, we propose a psychoacoustic ANC system which considers the spectral modulation sensitivity of human hearing to shape the spectrum of the residual error. The concept of considering different hearing thresholds of acoustic frequencies in developing psychoacoustic ANC systems [4][5] is extended to include different detection thresholds of spectral modulations. We extend the hybrid ANC system in [5] by further incorporating another pre-emphasis filter, which accounts for the spectral modulation sensitivity of human hearing.

The organization of this paper is as follows. In section 2, the perceptual effect due to different sensitivities to spectral modulations is described. In section 3, we describe the proposed psychoacoustic ANC system with the additional pre-emphasis filter. The experimental results of objective and subjective tests are presented in section 4. Finally, we end in section 5 with conclusions and future work.

2. Spectral Modulation Sensitivity

In this section, a brief introduction of an auditory model is given to demonstrate the perceptual effect due to different detection thresholds of spectral modulations. This auditory model consists of two parts: (1) an early cochlear module, which transforms the input acoustic signal into an auditory spectrogram, and (2) a central cortical module which further analyzes the auditory spectrogram in terms of the joint spectro-temporal modulations. Detailed descriptions of this auditory model can be accessed in [6]. In our approach, the noise signal was first processed by the cochlear module and then weighted by human hearing sensitivities to spectral modulations. In this way, we can construct a static noise-shaping filter for the FELMS ANC system to put more emphasis on critical spectral modulation bands while cancelling the noise.

2.1. Cochlear module

The early cochlear module transforms the input acoustic signal from a time-domain waveform into an auditory spectrogram by incorporating critical neural activities along the peripheral auditory pathway. The cochlear module first filters the input sound using 128 constant-Q band-pass filters which simulate the quasi-logarithmic frequency selectivity of the cochlea. The

  http://dx.doi.org/10.21437/Interspeech.2016-757

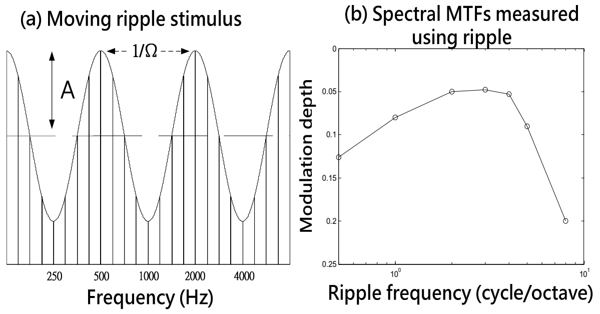

Figure 1: *(a) Spectral profile of a moving ripple stimulus; (b) detection thresholds of spectral modulations [11].*

output of each filter is then passed through a non-linear compression stage and a lateral inhibitory network (LIN). The non-linear compression stage models the saturation of the inner hair cells, and the LIN models the frequency masking effect. Outputs of the LIN are further processed by a half-wave rectifier followed by a low-pass filter to extract envelopes and form an auditory spectrogram. The auditory spectrogram records the envelope profile of a speech signal in the joint time and log-frequency domain unlike the joint time and linear frequency domain of the Fourier transform based spectrogram. More detailed descriptions and formulations of this cochlear module can be found in [6].

2.2. Spectral modulation sensitivity

Spectro-temporal modulations have been shown critical to speech intelligibility and speech quality [7][9]. However, for building an ANC with static pre-emphasis filters, only the spectral modulation sensitivity of human hearing is considered in this study. The sensitivity of the auditory cortex to spectral modulations was studied using ripple stimuli [11]. Each ripple stimulus was constructed using equally spaced frequency components on the logarithmic frequency axis. The ripple stimuli can be thought as eigenfunctions of the auditory cortex, which was assumed a quasi-linear system [11]. As shown in Figure 1(a), the spectral profile S(x) of each ripple stimulus can be written as:

$$S(x) = A \cdot \sin(2 \cdot \pi \cdot \Omega \cdot x + \Phi) \qquad (1)$$

where A is defined as the modulation depth (ranging from 0 to 1); x represents the logarithmic frequency; Ω (scale) is defined as the ripple density in units of cycle/octave (cyc/oct); and the phase Φ determines the starting position of the sine function. The measured detection thresholds of spectral modulations with Ω = 0.5, 1, 2, 3, 4, 5 and 8 cyc/oct were shown in [11] and is replotted in Figure 1(b). This curve depicts the spectral modulation sensitivity and is called the spectral modulation transfer function (SMTF). This low-pass shaped SMTF indicates test subjects were more sensitive to 1~4 cyc/oct spectral modulations and lost their ability to detect rapid changing spectral profiles (Ω > 5).

The spectral modulation gain function (SMGF) was derived by inverting the detection thresholds in Figure 1(b), i.e., a higher/lower gain is assigned to the modulation band with a lower/higher threshold. The SMGF can be thought as the gain function of the brain to spectral modulations. Figure 2(a) shows a normalized Fourier spectral profile (magnitude spectrum) of an arbitrary frame of white noise. Figure 2(b) shows the corresponding weighted spectral profile after

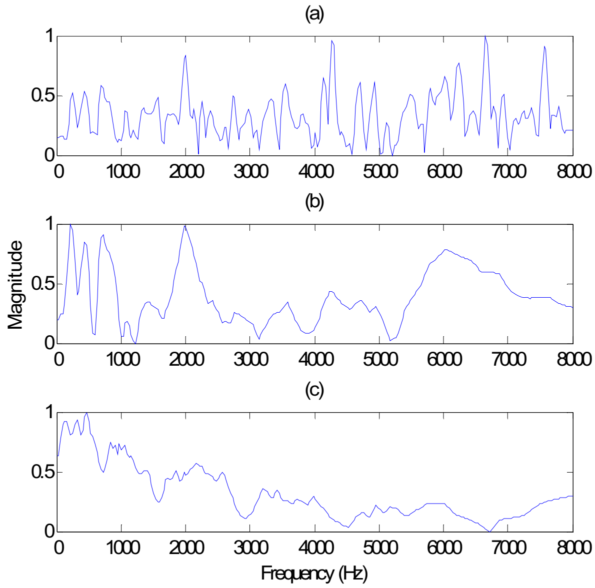

Figure 2: *(a) An arbitrary Fourier magnitude spectrum of white noise at a given time instant; (b) the corresponding SMGF-weighted spectrum; (c) the mean of SMGF-weighted Fourier magnitude spectra of white noise over 500ms.*

applying the SMGF. The magnitude response at high frequencies (> 2 kHz) is smoothed and suppressed because the high-frequency peaks constitute high-scale spectral modulations (more peaks per octave), which are suppressed by the SMGF. Figure 2(c) shows the averaged SMGF-weighted Fourier magnitude spectrum of white noise over 500 ms. This figure demonstrates human hearing is more sensitive to frequency fluctuations of white noise at lower frequencies than at higher frequencies due to different sensitivities to spectral modulations.

2.3. SMGF-weighted pre-emphasis filter

In [5], a pre-emphasis filter was roughly generated by applying A-weighting [12] which mimics the acoustic frequency gain function (AFGF) of human hearing. In this paper, we aim to derive the second pre-emphasis filter to account for the SMGF of human hearing. To fully grasp SMGF effects, a time-varying pre-emphasis filter is needed and can be constructed by applying the SMGF to the input spectra of the noise signal frame by frame. However, it is impractical to implement such a dynamic pre-emphasis filter in an ANC system for real-time process because the computational cost of applying the SMGF frame by frame is too high.

Therefore, in our system, offline modeling was used for deriving the second static pre-emphasis filter. The SMGF weighted Fourier spectrum of each frame of white noise was produced and then averaged over 500ms to generate a representative spectral profile as shown in Figure 2(c). Finally, a smoothed version of the averaged profile was adopted as the second pre-emphasis filter as shown in Figure 3.

3. Proposed ANC System

Figure 4 shows the block diagram of the psychoacoustic hybrid ANC system in [5], which combines the FELMS

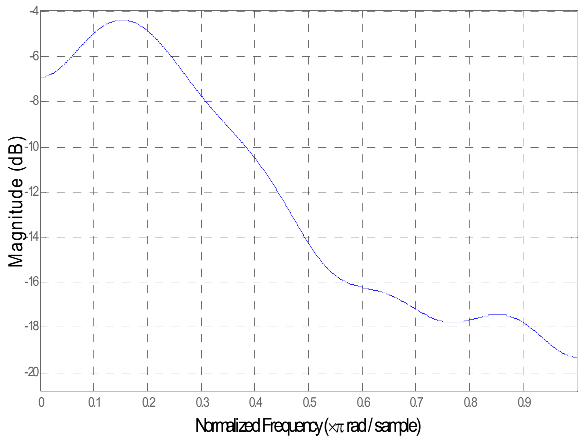

Figure 3: *Frequency response of the second pre-emphasis filter for white noise.*

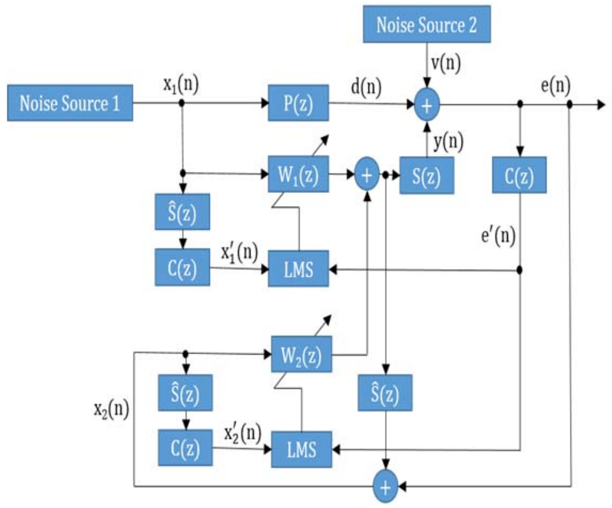

Figure 4: *Block diagram of the psychoacoustic hybrid ANC system.*

algorithm with the typical hybrid ANC system. In this paper, we extend this block diagram by adding the second pre-emphasis filter to take the SMGF of human hearing into account. The update rules of the coefficients of the two adaptive filters $W_1(z)$ and $W_2(z)$ can be directly extended from [5] as:

$$w_1(n+1) = w_1(n) + \mu_1 e'(n)[x_1(n) * \hat{s}(n) * c(n)] \quad (2)$$

$$w_2(n+1) = w_2(n) + \mu_2 e'(n)[x_2(n) * \hat{s}(n) * c(n)] \quad (3)$$

where $w_1(n)$ is the coefficient of the feedforward adaptive filter; $w_2(n)$ is the coefficient of the feedback adaptive filter; μ_1 and μ_2 are the step sizes; $e'(n)$ is the filtered error signal; $c(n)$ is the overall impulse response of two pre-emphasis filters, including the AFGF filter and the proposed SMGF filter. These two adaptive filters are used to estimate an unknown primary path $P(z)$, which accounts for the response of the acoustic path from the reference microphone to the error microphone. The overall electrical and acoustic path from the output of these adaptive filters to $e(n)$ is described by the secondary path $S(z)$ which includes system functions of the digital-to-analog (D/A) converter, the reconstruction filter, the power amplifier, the loudspeaker, the acoustic path from loudspeaker to the error microphone, the error microphone, the pre-amplifier, the antialiasing filter, and the analog-to-digital (A/D) converter. The $\hat{s}(z)$, the estimate of $S(z)$, can be obtained using either on-line or off-line techniques [1].

In addition to the conventional SPL measure, the objective loudness measure was also used as in [4][5] to assess performance of the proposed ANC system. The objective loudness measure can be written as:

$$L = \int_0^{24Bark} N\,dz \quad (4)$$

where L is the loudness and N is the loudness in a specific critical band in the unit of sone/Bark. Bark is the scale where the critical bandwidth of hearing is defined [13].

4. Simulation Results

In this section, the proposed system is evaluated and compared with the conventional hybrid ANC system [1] and the psychoacoustic hybrid ANC system in [5] for cancelling the narrowband and broadband noise. The SPL and loudness level of the residual noise were measured for comparison after each ANC system converged.

In our simulations, the primary path transfer function $P(z)$ and the secondary path transfer function $S(z)$ were derived using data provided by the MIT Media Laboratory [14]. Both impulse responses were truncated to 90 points. For the case of narrowband noise, the noise source $x_1(n)$ was a periodic signal comprised of five tones of frequencies 0.2, 0.3, 1, 2, and 3 kHz. The uncorrelated noise source $v(n)$ comprised of another five tones of frequencies 0.15, 0.25, 0.9, 1.5, and 2.5 kHz. This kind of multi-tone test was often seen in ANC literature [4][5]. For the case of broadband noise, two kinds of noise, factory and Volvo, were extracted from NOISEX-92 dataset for test. The two adaptive filters $W_1(z)$ and $W_2(z)$ had the same length of 32 taps in both narrowband and broadband tests.

4.1. Experiment of narrowband noise

Test results for narrowband noise are shown in Table 1, where $e_h(n)$, $e_{ph}(n)$ and $e_{proposed}(n)$ are the residual noise of the conventional hybrid ANC system, of the psychoacoustic hybrid ANC system in [5], and of our proposed system, respectively. The $d(n) + v(n)$ is the total noise when the ANC system is turned off. The results show that the residual noise of the conventional hybrid ANC system has the lowest SPL. However, by adding the hearing-perception inspired pre-emphasis filters, the loudness of the residual noise can be reduced significantly. Comparing $e_{proposed}(n)$ with $e_h(n)$, although the SPL of the residual noise increases by 1.67 dB (2.4%), the loudness decreases by 20.9%. Comparing $e_{proposed}(n)$ with $e_{ph}(n)$, our proposed system achieves a 8.1% lower loudness level but with a 1.3% higher SPL than the psychoacoustic hybrid ANC system in [5]. The slight 1.3% increase of SPL of the proposed system is tolerable since human hearing is more sensitive to the loudness level than the SPL [4][5].

Figure 5 shows the loudness level and SPL of the residual noise of the three compared systems against time. It clearly shows three systems have very similar convergence rates. Figure 5(a) demonstrates that our proposed system outperforms other two systems in loudness measure after

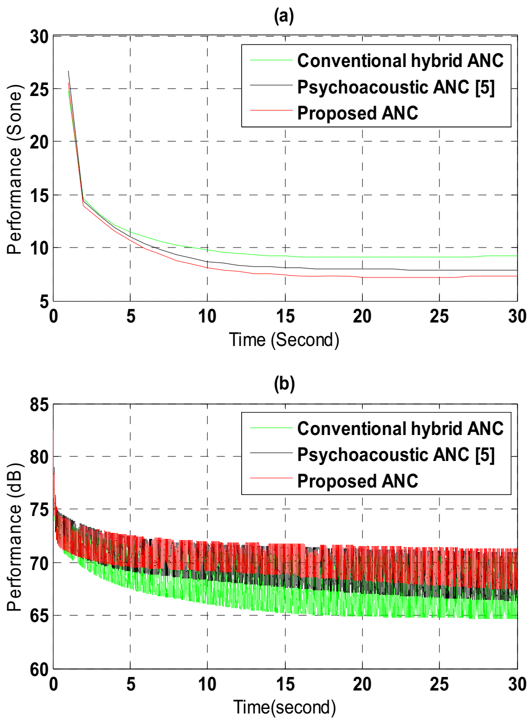

Figure 5: *Noise reduction performance in terms of* (*a*) *loudness (sone);* (*b*) *SPL (dB).*

around 2 seconds. Comparing our proposed system with the conventional hybrid ANC, our loudness level is about 0.5 to 0.8 sones lower from 2 to 5 seconds, and about 1 to 2 sones lower after 5 seconds. Comparing with the psychoacoustic hybrid ANC in [5], the loudness level of our proposed ANC is about 0.4 to 1 sone lower after 2 seconds.

In addition to the objective SPL and loudness measures, subjective listening tests were also conducted for evaluations. Sixteen subjects aged between 22 and 28 were recruited for the two-alternative preference test [15]. In the tests, subjects were asked to tell which residual noise was more perceivable. During the tests, the sequence of test signals was randomly assigned and the signals can be played repeatedly. Two listening tests were conducted: (a) the conventional hybrid ANC versus the proposed ANC; and (b) the psychoacoustic hybrid ANC in [5] versus the proposed ANC. The listening test results are given in Table 2, where n_i = {1, 3, 5, 8, 28} are the test time instants in seconds. The numbers in Table 2 are the numbers of subjects who reported the residual noise of the conventional hybrid ANC or of the psychoacoustic hybrid ANC in [5] was more perceivable than the residual noise of our proposed ANC.

Table 1: *Objective measures of the residual noise of three compared ANC systems in narrowband noise*

	SPL(dB)	Loudness(sone)
d(n) + v(n)	79.1813	34.4100
$e_h(n)$	67.7382	9.1910
$e_{ph}(n)$	68.5525	7.9180
$e_{proposed}(n)$	69.4116	7.2750

Table 2: *Subjective preference listening test results for narrowband noise*

Test	Time Index n_1	n_2	n_3	n_4	n_5
(a)	12	10	13	14	15
(b)	13	9	11	12	13

Table 3: *Objective measures of the residual noise of three compared ANC systems in two kinds of broadband noise*

	factory		Volvo	
	SPL(dB)	Loudness (sone)	SPL(dB)	Loudness (sone)
d(n)	82.3649	29.21	85.7778	19.91
$e_h(n)$	69.9996	11.097	78.8101	15.429
$e_{ph}(n)$	69.4713	9.132	79.1191	13.929
$e_{proposed}(n)$	70.0246	7.784	79.7511	7.763

These results clearly show the residual noise of both the conventional hybrid ANC and the psychoacoustic hybrid ANC system in [5] is more perceivable than the residual noise of our proposed ANC at all time instants especially after n_3. From the objective loudness measure and subjective listening test results, we conclude our proposed system produces less perceivable residual noise than the other two compared systems.

4.2. Experiment of broadband noise

The factory and Volvo car noise were used for the broadband noise test. Due to the non-stationarity of the noise, we calculated the mean SPL and loudness level of the residual noise of the three tested ANC systems by averaging over three seconds after the systems converged. Results are shown in Table 3. These results clearly show that our proposed ANC system produces the residual noise with the lowest loudness level comparing with other two ANC systems. Subjective preference listening tests were also conducted for broadband noise. Results (not shown in this paper) also demonstrate the proposed system produces less perceivable residual noise than the other two compared systems.

5. Conclusions

In this paper, we aim to improve the performance of ANC systems by considering spectral modulation sensitivity of human hearing. Simulation results demonstrate that the proposed system produces better results in objective and subjective tests. The price paid for this enhanced performance is the inclusion of an additional static pre-emphasis filter. We have validated the proposed method on a TI C6713 DSP platform for real-time process. In the future, we will collaborate with industrial companies to prototype the system for real-world test.

6. Acknowledgements

This research is supported by the Ministry of Science and Technology, Taiwan under Grant No MOST 103-2220-E-009-003.

7. References

[1] S. M. Kuo and D. R. Morgan, "Active noise control: A tutorial review," in *Proc. IEEE*, vol. 8, no. 6, pp. 943-973, Jun. 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, doi: 10.1017/ATSIP.2012.4, 2012.

[3] S. M. Kuo and J. Tsai, "Residual noise shaping technique for active noise control systems," *J. Acoust. Soc. Am.*, vol. 95, no. 3, pp. 1665-1668, Mar. 1994.

[4] H. Bao and I. Panahi, "Psychoacoustic active noise control based on delayless subband adaptive filtering," in *Proc. ICASSP*, pp. 341-344, 2010.

[5] T. W. Wang, W. S. Gan and Y. K. Chong, "Psychoacoustic hybrid active noise control system," in *Proc. ICASSP*, pp. 321-324, 2012.

[6] T. Chi, P. Ru, and S. A. Shamma, "Multi-resolution spectro-temporal analysis of complex sounds," *J. Acoust. Soc. Am.*, vol. 118, no. 2, pp. 887-906, Aug. 2005.

[7] C.-C. Hsu, T.-E. Lin, J.-H. Chen and T.-S. Chi, "Spectro-temporal subband Wiener filter for speech enhancement," in *Proc. ICASSP*, pp. 4001-4004, 2012.

[8] C.-C. Hsu, T.-E. Lin, J.-H. Chen and T.-S. Chi, "Voice activity detection based on frequency modulation of harmonics," in *Proc. ICASSP*, pp. 6679-6683, 2013.

[9] M. Elhilali, T. Chi and S. Shamma, "A spectro-temporal modulation index (STMI) for assessment of speech intelligibility," *Speech Communication*, pp. 331-348, Oct. 2003.

[10] T. Chi, Y. Gao, M. C. Huyton, P. Ru and S. A. Shamma, "Spectro-temporal modulation transfer functions and speech intelligibility," *J. Acoust. Soc. Am.*, vol. 106, no. 5, pp. 2719-2738, Nov 1999.

[11] D. Green, "Frequency and the detection of spectral shape change," in *Auditory Frequency Selectivity*, edited by B. C. J. Moore and R. Patterson, Plenum, Cambridge, pp. 351-359, 1986.

[12] *ANSI* Standards S1.42-2001.

[13] *ISO 532B*, "Acoustics – Method for calculating loudness level," International Organization for Standardization, 1975.

[14] B. Gardner, K. Martin, *HRTF measurements of a KEMAR dummy-head microphone,* MIT Media Lab Perceptual Computing Technical Report #280, 1994.

[15] E. Zwicker, *Psychoacoustics: Facts and Models, 2nd ed.* New York: Springer, 1999.