ON THE USE OF A PHASE MODULATION METHOD FOR DECORRELATION IN ACOUSTIC FEEDBACK CANCELLATION

Meng Guo[1,2], *Søren Holdt Jensen*[2], *and Jesper Jensen*[1], *and Steven L. Grant*[3*]

[1]Oticon A/S, Kongebakken 9, DK-2765 Smørum, Denmark.
[2]Department of Electronic Systems,
Aalborg University, Niels Jernes Vej 12, DK-9220 Aalborg, Denmark.
[3]Department of Electrical and Computer Engineering,
Missouri University of Science and Technology, 301 West 16th Street, Rolla, MO 65409, USA.
emails: {guo, jsj}@oticon.dk, shj@es.aau.dk, sgrant@mst.edu

ABSTRACT

A major problem in using an adaptive filter in acoustic feedback cancellation systems is that the loudspeaker signal is correlated with the signals entering the microphones of the audio system, leading to biased filter estimates. One possible solution for reducing this problem is by means of decorrelation. In this work, we study a subband phase modulation method, which was originally proposed for decorrelation in multichannel acoustic echo cancellation systems. We determine if this method is effective for decorrelation in acoustic feedback cancellation systems by comparing it to a structurally similar frequency shifting decorrelation method. We show that the phase modulation method is suitable for decorrelation in a hearing aid acoustic feedback cancellation system, although the frequency shifting method is in general slightly more effective.

Index Terms— Acoustic feedback cancellation, adaptive filters, decorrelation, phase modulation, frequency shifting.

1. INTRODUCTION

Adaptive filters have been widely used in both acoustic echo cancellation (AEC) for audio and communication systems and acoustic feedback cancellation (AFC) for sound reinforcement systems. The goal of the adaptive filters in both cases is to model the acoustic signal paths from loudspeakers to microphones of audio systems.

A major problem when using adaptive filters in stereo and/or multichannel AEC systems is the so-called non-uniqueness problem due to the fact that the loudspeaker signals are strongly correlated [1]. It can be shown that the adaptive filter estimates do not converge correctly to the true acoustic echo paths. In AFC systems, on the other hand, the main problem in using adaptive filters is the biased adaptive filter estimation of the acoustic feedback paths [2], which is caused by the nonzero correlation between the loudspeaker signals and the signals entering the microphones.

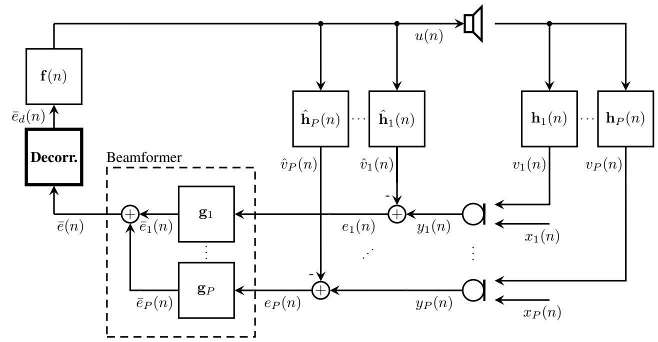

Fig. 1. A general AFC system with a decorrelation function.

In both cases, the biased filter estimation is due to undesired and unavoidable signal correlations in audio systems, although the causes of these signal correlations are different. Many decorrelation methods have been proposed for both stereo AEC and AFC systems in the past. A simple method is to introduce nonlinear distortions to loudspeaker signals as firstly proposed for stereo AEC systems [3] and later studied for AFC systems [4]. Another widely used decorrelation method for both stereo AEC and AFC systems is performed by adding uncorrelated noise to the loudspeaker signals; the added noise is preferably generated such that it is inaudible in the presence of the loudspeaker signals, see e.g. [5, 6, 7]. Some other proposed decorrelation methods include introducing time-variable delays on the loudspeaker signals [8], using variable all-pass filtering on the loudspeaker signals to introduce phase shifts [9], applying decorrelation prefilters to the signals used for the adaptive filter estimation [10], and using frequency shifting of the loudspeaker signals [11]. Generally, all these methods might introduce sound quality degradations. Thus, an important compromise in using these methods is sound quality versus decorrelation ability and thereby cancellation performance improvement.

In this work, we study decorrelation methods in an AFC

*This work was funded in part under the Wilkens Missouri Endowment.

system as shown in Fig. 1, where the AFC is carried out by adaptive filters $\hat{\mathbf{h}}_i(n)$, where n is the time index, $i = 1, ..., P$, and P is the number of microphones. The goal of $\hat{\mathbf{h}}_i(n)$ is to cancel the effects of the true acoustic feedback paths $\mathbf{h}_i(n)$. Furthermore, beamformer filters $\mathbf{g}_i$ are performing a spatial filtering on the feedback compensated signals $e_i(n)$. The block "Decorr." denotes the applied decorrelation function, and the decorrelated signal $\bar{e}_d(n)$ is modified by the forward path $\mathbf{f}(n)$ to form the loudspeaker signal $u(n)$.

More specifically, we study a perceptually motivated decorrelation method by means of subband phase modulation for AFC systems. This phase modulation method was originally introduced for stereo and multichannel AEC systems in [12], where it provided a good cancellation performance without significant sound quality degradation. Here we will determine if this phase modulation method is useful for AFC systems, since not every decorrelation method suitable for AEC would necessarily be appropriate for AFC systems [4].

This phase modulation method is in structure very similar to a frequency shifting decorrelation method [4] carried out in a subband implementation. Thus, we find it obvious to compare both methods. Since the frequency shifting has already been evaluated among other decorrelation methods for AFC systems [4], this comparison will also reveal the effectiveness of the phase modulation method. In particular, we determine analytically the differences between these two methods, before we evaluate AFC performance by simulations, given that sound quality distortions are at same levels for both methods.

2. ANALYSIS OF DECORRELATION METHODS

In this section, we provide details on the subband phase modulation and frequency shifting decorrelation methods. Furthermore, we discuss the differences between them.

Both decorrelation methods are carried out in filter bank subbands, as shown in Fig. 2. An over-sampled analysis filter bank with a decimation factor D divides the input signal $w(n)$ into M subbands with subband index $m = 0, 1, \ldots, M-1$. A complex exponential function $e^{j\varphi(k,m)}$ is then multiplied on each filter bank subband signal $w(k, m)$ to create $z(k, m) = w(k, m)e^{j\varphi(k,m)}$, where $k = 0, 1, 2, \ldots$ is the subband time index with the corresponding fullband time index $n = 0, D, 2D, \ldots$. A synthesis filter bank recombines the processed subband signals $z(k, m)$ to a fullband signal $z(n)$. The difference between these two decorrelation methods is the choice of complex exponential functions $e^{j\varphi(k,m)}$.

The subband structure allows the use of different phase functions $\varphi(k, m)$ over subbands, and $\varphi(k, m)$ can be chosen based on human auditory perception to minimize sound quality degradation. In the phase modulation method proposed in [12], a smooth phase function $\varphi_p(t, m)$ to provide decorrelation with minor sound distortions was suggested as

$$\varphi_p(t, m) = \alpha(m) \sin(2\pi f_m t), \qquad (1)$$

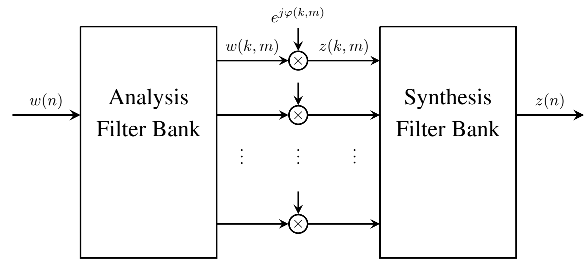

Fig. 2. Decorrelation with subband phase modifications.

where t denotes continuous time, $\alpha(m)$ is the phase amplitude for the mth subband, and f_m is the modulation frequency. In [12], the optimal values of $\alpha(m)$ and f_m were found by a listening procedure, so that effects of $\varphi_p(t, m)$ would be perceptually insignificant. In particular, a modulation frequency $f_m = 0.75$ Hz was suggested, whereas the phase amplitudes $\alpha(m)$ varied from 10 degrees at low frequencies to 90 degrees above 2.5 kHz. Furthermore, complex conjugated phase functions $\varphi_p(t, m)$ were applied on both microphone channels in a stereo system.

Eq. (1) can also be expressed in the subband time index k using $t = \frac{k}{f_s/D}$, where f_s is the fullband sampling rate, as

$$\varphi_p(k, m) = \alpha(m) \sin\left(2\pi f_m k \frac{D}{f_s}\right). \qquad (2)$$

On the other hand, the frequency shifting method is carried out using the complex exponential function $e^{j\varphi_f(k,m)}$ in filter bank subbands, and the phase function $\varphi_f(k, m)$ is given by

$$\varphi_f(k, m) = 2\pi f_0(m) k \frac{D}{f_s}, \qquad (3)$$

where $f_0(m)$ denotes the amount of frequency shifting in subband m. Thus, for $f_0(m) > 0$, $\varphi_f(k, m)$ increases linearly with increasing k. Furthermore, using the modulus operator mod(), the wrapped version of $\varphi_f(k, m) > 0$ is expressed by

$$\varphi'_f(k, m) = \text{mod}\left(2\pi f_0(m) k \frac{D}{f_s}, 2\pi\right) - c, \qquad (4)$$

where $c = 0$ if $\text{mod}(2\pi f_0(m) k \frac{D}{f_s}, 2\pi) \leq \pi$, and $c = 2\pi$, otherwise.

Fig. 3 shows the phase functions $\varphi_p(k, m)$ and $\varphi'_f(k, m)$ given by Eqs. (2) and (4) for the phase modulation and frequency shifting methods, respectively. The time is computed as $t = \frac{k}{f_s/D}$. There are some obvious similarities, both functions are periodic with a certain frequency and amplitude. In the phase modulation case, the frequency and amplitude of $\varphi_p(k, m)$ are determined by f_m and $\alpha(m)$, respectively. In the frequency shifting case, the frequency of $\varphi'_f(k, m)$ is determined by $f_0(m)$, whereas the maximum and minimum amplitude values are always $\pm\pi$.

An important difference between these phase functions is that $\varphi'_f(k, m)$ is an increasing function in its unwrapped

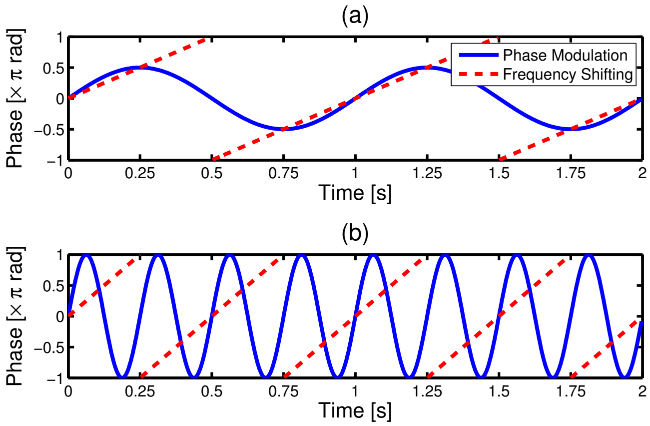

Fig. 3. Phase functions $\varphi_p(k,m)$ and $\varphi'_f(k,m)$ with different parameter values. (a) $\alpha(m) = \frac{\pi}{2}$, $f_m = 1$ Hz, and $f_0(m) = 1$ Hz. (b) $\alpha(m) = \pi$, $f_m = 4$ Hz, and $f_0(m) = 2$ Hz.

form $\varphi_f(k,m)$ given by Eq. (3), whereas $\varphi_p(k,m)$ is identical to its wrapped function and it is always periodic. Since a frequency shift is introduced proportionally to the temporal derivative of the phase function $\varphi(k,m)$, a constant frequency shift is obtained by using the phase function $\varphi_f(k,m)$, whereas with $\varphi_p(k,m)$ the amount of frequency shift is time-varying with an average of zero.

3. SOUND QUALITY CONSIDERATIONS

In this section, we choose parameters $f_0(m)$, $\alpha(m)$ and f_m for both decorrelation methods, so that they only introduce insignificant and somewhat equal sound quality degradations.

In many applications, a fullband frequency shifting factor f_0 is commonly chosen as $0 < f_0 \leq 10$ Hz to avoid significant sound quality distortions, as e.g. suggested in [4]. In this work, we let f_0 be subband dependent as $f_0(m)$, and we only perform frequency shifting at higher frequencies to further preserve sound quality, especially for speech signals. We use filter banks with $M = 64$ complex conjugated subbands with a fullband sampling rate of $f_s = 20$ kHz and a decimation factor $D = 8$. Table 1 shows the chosen shifting factors $f_0(m)$ for subbands $m = 0, \ldots, 31$.

The phase function $\varphi_p(k,m)$ for the phase modulation method has two parameters, $\alpha(m)$ and f_m. In this work, we use the phase amplitude values $\alpha(m)$ as given in Table 2. We chose these amplitude values to match the introduced phase amplitude differences between microphone channels in [12]. Furthermore, as demonstrated in Fig. 3, the phase modulation frequency f_m determines the frequency of the periodic phase function $\varphi_p(k,m)$ as the shifting frequency $f_0(m)$ does for $\varphi'_f(k,m)$. From this relation, we choose $f_m = 10$ Hz in an attempt to obtain similar sound qualities in both methods.

We now perform objective sound quality measurements to verify our parameter choices, by using the MATLAB implementations [13, 14] of perceptual evaluation of speech quality

Table 1. The frequency shifting factor $f_0(m)$ for subbands $m = 0, \ldots, 31$. The subband bandwidth is 312.5 Hz.

Subband m	0	1	2	3	≥ 4
$f_0(m)$ **[Hz]**	0	0	0	0	10

Table 2. The phase amplitude $\alpha(m)$ for subbands $m = 0, \ldots, 31$. The subband bandwidth is 312.5 Hz.

Subband m	$0-3$	4	5	6	≥ 7
$\alpha(m)$ **[Degrees]**	20	40	70	90	180

Table 3. The output scores from PESQ and PEAQ models.

PESQ	**PEAQ**	**Sound Quality/Description**
4.5	0	Excellent/Imperceptible
4	-1	Good/Perceptible but not annoying
3	-2	Fair/Slightly annoying
2	-3	Poor/Annoying
1	-4	Bad/Very annoying

Table 4. The mean, standard deviation, and median values of the PESQ and PEAQ scores based on 9 speech signals and 9 music signals processed by the frequency shifting (**FS**) and phase modulation (**PM**) decorrelation methods.

Statistics	**PESQ**		**PEAQ**	
	FS	**PM**	**FS**	**PM**
Mean	4.26	4.30	-1.29	-1.34
Stdv.	0.04	0.11	0.69	0.59
Median	4.27	4.29	-1.00	-1.05

(PESQ) and perceptual evaluation of audio quality (PEAQ) models, described in [15] and [16], respectively. Table 3 provides descriptions of the output scores from both models.

Although both models were originally developed to assess relatively mild coding artifacts, we use them to evaluate the relatively small degradations from both decorrelation methods. We compare several test signals $w(n)$ with their processed versions $z(n)$ obtained as shown in Fig. 2 using both decorrelation methods. Table 4 shows the mean, standard deviation, and median values of the determined sound quality scores based on a total of 18 speech and music test signals.

The quality scores in Table 4 show that similar sound quality degradations can be expected from both decorrelation methods, with the chosen parameters of $f_0(m)$, $\alpha(m)$ and f_m. Furthermore, the sound quality degradations are limited in both cases, especially for speech signals. From the determined PESQ scores we can classify the degradation as only slightly perceptual but not annoying. This is because none or only minor modifications are carried out in frequency regions below approximately 1.5 kHz. For music signals, however, the sound quality degradations are more severe since they generally have more high frequency contents. Nevertheless, the introduced degradations can still be roughly characterized as perceptible but not annoying. Furthermore, the result from the objective sound quality evaluation was also confirmed by a few experienced listeners.

4. SIMULATION EXPERIMENTS

In this section, we perform simulations to evaluate both decorrelation methods in a hearing aid AFC system as shown in Fig. 1 with two microphones ($P = 2$). We determine how effective the phase modulation method is compared to the frequency shifting method, when both methods have sound quality degradations at similar levels as determined in Sec. 3.

The true acoustic feedback paths $\mathbf{h}_i(n)$ remain time invariant in simulations, i.e. $\mathbf{h}_i(n) = \mathbf{h}_i$, and they are obtained by measurements from a behind-the-ear hearing aid with two microphones. Fig. 4 shows the impulse responses $\mathbf{h}_i$ for both microphone channels, where the sampling rate is 20 kHz. From Fig. 4 we observe that the effective length, which covers the nonzero values of $\mathbf{h}_i$, is approximately 50 taps.

The feedback path estimates $\hat{\mathbf{h}}_i(n)$ are obtained using a delayless subband adaptive filter approach [17, 18] with 32 subbands and a decimation factor of 16, where the corresponding fullband adaptive filters $\hat{\mathbf{h}}_i(n)$ have 64 taps, and the subband NLMS algorithm utilizes a step size of $\mu = 2^{-11}$ for subbands above approximately 1.25 kHz, and $\mu = 0$ for low frequency subbands. This is motivated by the fact that AFC is usually not necessary for the lowest frequencies in a hearing aid application. Furthermore, the regularization parameter $\delta = 2^{-14}$ is used in the subband NLMS algorithm.

We evaluate the AFC performance using the coefficient misalignment criterion $\epsilon(n)$ defined as

$$\epsilon(n) = \frac{\sum_{i=1}^{P} \|\mathbf{h}_i - \hat{\mathbf{h}}_i(n)\|^2}{\sum_{i=1}^{P} \|\mathbf{h}_i\|^2}. \quad (5)$$

In simulations, the incoming signal $x_1(n)$ is either a speech or music signal, and we simply use a delay of one sample between $x_2(n)$ and $x_1(n)$ to model the distance between the microphones. This delay corresponds to a microphone distance of approximately 17 mm, assuming the microphones are aligned in front/rear positions in a horizontal plane, and the sound signal is coming from the front direction.

We use a simple beamformer setup as $\mathbf{g}_1 = \mathbf{g}_2 = \frac{1}{2}$, whereas the forward path $\mathbf{f}(n)$ consists of a delay of 120 samples, corresponding to a hearing aid processing delay of 6 ms. Furthermore, $\mathbf{f}(n)$ consists of a single-channel fullband compressor to provide a maximum amplification of 29.3 dB, and the most critical closed-loop magnitude value without $\hat{\mathbf{h}}_i(n)$ becomes -1 dB at approximately 2.5 kHz.

We compare simulation results in terms of $\epsilon(n)$ between a reference AFC system without applying any decorrelation method and two AFC systems using each decorrelation

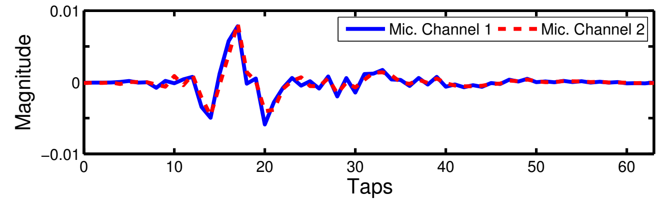

Fig. 4. Measured acoustic feedback paths from a behind-the-ear hearing aid with a sampling rate of $f_s = 20$ kHz.

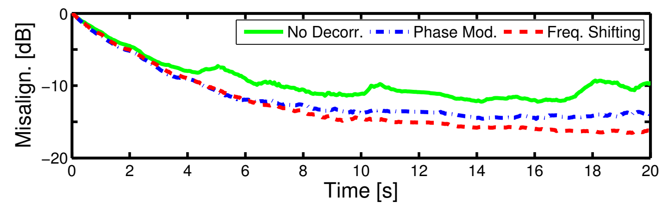

Fig. 5. A representative example simulation result showing the misalignments $\epsilon(n)$ for three different AFC systems.

Table 5. Simple statistics for steady-state misalignments [dB] in AFC systems without decorrelation (**None**), with phase modulation (**PM**) and frequency shifting (**FS**), respectively.

Stat.	**Speech**			**Music**		
	None	**PM**	**FS**	**None**	**PM**	**FS**
Mean	-14.6	-16.0	-16.0	-7.6	-13.6	-14.8
Stdv.	1.4	0.5	0.9	3.6	2.4	2.7
Median	-14.3	-16.0	-16.2	-9.2	-14.5	-16.2

method. We performed many simulation trials with different speech and music signals as the incoming signals $x_i(n)$. Fig. 5 shows a representative example result for a music signal. We observe that a smaller misalignment $\epsilon(n)$ is already obtained after about 1 s, by using either decorrelation method compared to the system without decorrelation, and the improvement is more than 6 dB in the frequency shifting case at the end of the simulation.

Table 5 shows statistics of the steady-state misalignments from all simulations. For speech signals, the steady-state misalignments $\epsilon(n)$ are close (within 1.4 dB) between all three systems. In fact, $\epsilon(n)$ generally followed each other closely over time (not shown). Hence, there is only small improvements by using either decorrelation method in this case. This is because speech signals in general only cause a limited correlation problem for the estimation of $\hat{\mathbf{h}}_i(n)$ at the higher frequencies above approximately 1.5 kHz, so a decorrelation in this region can not improve AFC performance much further.

On the other hand, the correlation problem is generally more severe for most music signals in the frequency region above 1.5 kHz, and we can expect a more significant AFC improvement when using decorrelation. This is confirmed by our simulations. In Table 5, we observe the average steady-state misalignment improvements for music signals are 6 dB and 7.2 dB in the phase modulation and frequency shifting cases, respectively. In some simulations, these improvements were found to be more than 12 dB. Furthermore, Table 5 also reveals that similar AFC performance to speech signals can be achieved for music signals when using decorrelation.

We should emphasize that the AFC performance improvements achieved from both decorrelation methods depend on the compromise made in sound quality. Thus, we conclude based on the chosen parameter values and sound quality evaluations done in Sec. 3, that the phase modulation decorrelation method is effective in improving hearing aid AFC performance, especially for music signals. However, the structurally very similar frequency shifting decorrelation method is generally slightly better for doing so.

5. CONCLUSION

In this work, we studied a subband phase modulation decorrelation method originally proposed for stereo and multichannel AEC systems. We compared it to a similar frequency shifting decorrelation method in an AFC system. We determined analytically the differences between these two decorrelation methods. Furthermore, we showed that by choosing appropriate parameter values in the phase modulation method, it is capable of improving the AFC performance without introducing significant sound quality degradations. However, by controlling sound quality degradations at similar and insignificant levels in both methods, the frequency shifting decorrelation method gives slightly better overall AFC performance, which probably makes it the preferred method, especially when taking its simplicity into account.

6. REFERENCES

[1] M. M. Sondhi, D. R. Morgan, and J. L. Hall, "Stereophonic acoustic echo cancellation - an overview of the fundamental problem," *IEEE Signal Process. Lett.*, vol. 2, pp. 148–151, 1996.

[2] A. Spriet, G. Rombouts, M. Moonen, and J. Wouters, "Adaptive feedback cancellation in hearing aids," *J. Franklin Inst.*, vol. 343, no. 6, pp. 545 – 573, 2006.

[3] J. Benesty, M. M. Sondhi, and D. R. Morgan, "A better understanding and an improved solution to the specific problems of stereophonic acoustic echo cancellation," *IEEE Trans. Speech Audio Process.*, vol. 6, pp. 156–165, 1998.

[4] T. van Waterschoot and M. Moonen, "Assessing the acoustic feedback control performance of adaptive feedback cancellation in sound reinforcement systems," in *Proc. 17th Eur. Signal Process. Conf.*, pp. 1997–2001, 2009.

[5] T. Gänsler and P. Eneroth, "Influence of audio coding on stereophonic acoustic echo cancellation," in *Proc. 1998 IEEE Int. Conf. Acoust., Speech, Signal Process.*, pp. 3649–3652, 1998.

[6] A. Gilloire and V. Turbin, "Using auditory properties to improve the behaviour of stereophonic acoustic echo cancellers," in *Proc. 1998 IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 6, pp. 3681–3684, 1998.

[7] M. Guo, S. H. Jensen, and J. Jensen, "An improved probe noise approach for acoustic feedback cancellation," in *Proc. 7th IEEE Sensor Array Multichannel Signal Process. Workshop*, pp. 505–508, 2012.

[8] A. Sugiyama, Y. Joncour, and A. Hirano, "A stereo echo cancellation with correct echo path identification based on an input-sliding technique," *IEEE Trans. Signal Process.*, vol. 49, no. 1, pp. 2577–2587, 2001.

[9] M. Ali, "Stereophonic acoustic echo cancellation system using time-varying all-pass filtering for signal decorrelation," in *Proc. 1998 IEEE Int. Conf. Acoust., Speech, Signal Process.*, pp. 3689–3692, 1998.

[10] A. Spriet, I. Proudler, M. Moonen, and J. Wouters, "Adaptive feedback cancellation in hearing aids with linear prediction of the desired signal," *IEEE Trans. Signal Process.*, vol. 53, no. 10, pp. 3749–3763, 2005.

[11] M. R. Schroeder, "Improvement of acoustic-feedback stability by frequency shifting," *J. Acoust. Soc. Am.*, vol. 36, no. 9, pp. 1718–1724, 1964.

[12] J. Herre, H. Buchner, and W. Kellermann, "Acoustic echo cancellation for surround sound using perceptually motivated convergence enhancement," in *Proc. 2007 IEEE Int. Conf. Acoust., Speech, Signal Process.*, pp. I–17–I–20.

[13] P. C. Loizou, *Speech Enahancement: Theory and Practice*. CRC Press, 2007.

[14] P. Kabal, *Perceptual Evaluation of Audio Quality (PEAQ)*. `http://www-mmsp.ece.mcgill.ca/documents/`, 2004.

[15] ITU-T Recommendation P.862, *Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs*. Int. Telecommun. Union, 2001.

[16] ITU-R Recommendation BS.1387-1, *Method for objective measurements of perceived audio quality*. Int. Telecommun. Union, 2001.

[17] D. R. Morgan and J. C. Thi, "A delayless subband adaptive filter architecture," *IEEE Trans. Signal Process.*, vol. 43, no. 8, pp. 1819–1830, 1995.

[18] J. Huo, S. Nordholm, and Z. Zang, "New weight transform schemes for delayless subband adaptive filtering," in *Proc. IEEE Global Telecommun. Conf.*, vol. 1, pp. 197–201, 2001.