Title
Speech quality and stable gain trade-offs in adaptive feedback cancellation for hearing aids.

Journal
The Journal of the Acoustical Society of America, 142(4)

ISSN
0001-4966

Authors
Lee, Ching-Hua
Kates, James M
Rao, Bhaskar D
et al.

Publication Date
2017-10-01

DOI
10.1121/1.5007278

Peer reviewed

Speech quality and stable gain trade-offs in adaptive feedback cancellation for hearing aids

Ching-Hua Lee,[1,a] James M. Kates,[2] Bhaskar D. Rao,[1] and Harinath Garudadri[1]

[1]*Department of Electrical and Computer Engineering, University of California, San Diego, California 92093, USA*

[2]*Department of Speech, Language, and Hearing Sciences, University of Colorado, Boulder, Colorado 80309, USA*

chl438@ucsd.edu, james.kates@colorado.edu, brao@ucsd.edu, hgarudadri@ucsd.edu

Abstract: This paper addresses trade-offs in adaptive feedback cancellation (AFC) for hearing aids. Aggressive AFC for improved added stable gain (ASG) reduces speech quality. In this paper, the hearing-aid speech quality index (HASQI) is used to investigate AFC performance before the system becomes unstable. It is demonstrated that for a desired speech quality, multiple AFC algorithms can be evaluated for their ASG and computational efficiency. An example is presented with HASQI = 0.8, baseline AFC, and two advanced approaches. For the advanced AFCs, ASG gains of 4 and 7 dB were obtained at additional computational complexity of 8% and 11%, respectively.

[CCC]

Date Received: June 30, 2017 **Date Accepted:** September 30, 2017

1. Introduction

In hearing aids (HAs), acoustic feedback is a well-known problem that causes howling and whistling effects which are annoying to the users. Under certain conditions, the receiver signal feeding back to the microphone will make the system become unstable. This not only degrades the audio quality of the output signal, but also limits the maximum stable gain (MSG) and the amount of amplification that can be provided by the HA device. To overcome this problem, many adaptive feedback cancellation (AFC) techniques have been proposed.[1,2] With the introduction of AFC, there is a trade-off between improved added stable gain (ASG) and reduced speech quality. The ASG is defined as the amount of additional MSG that can be provided by the AFC. In this paper, we investigate various AFC strategies to improve ASG for a desired speech quality. We propose to use an objective metric of speech quality as measured by the hearing-aid speech quality index (HASQI) version-2.[3] A real-world feedback path was used to provide the acoustic feedback.[4] The path response was measured using a microphone in a behind-the-ear HA case positioned behind the pinna on a dummy head, and the HA receiver was connected via tubing to the ear canal fitted with a vented earmold. Feedback path measurements were made with and without a telephone handset present. Speech and music signals from TIMIT and MPEG-4 test sequences are used to evaluate ASG for a target HASQI score.

Dynamic range compression is typically implemented in HAs. However, compression reduces the gain as the signal intensity increases, and can thus improve the system stability and reduce the severity of the acoustic feedback. In addition, dynamic range compression introduces nonlinear distortion and causes an associated reduction in the HASQI value, which would be present in addition to the effects of feedback and feedback cancellation. For the results reported below, we thus used linear amplification (i.e., compression ratio = 1) in order to focus on the feedback cancellation algorithm performance for the most challenging situation and to avoid the added confound of compression algorithm behavior. Therefore, a uniform gain over all the frequency bands was applied and varied to assess ASG for a given quality, as measured by HASQI.

As the gain of the HA increases, it becomes more difficult to cancel out the feedback signal, and may lower the quality of the processed signal, thus reducing HASQI. One question is: What is an acceptable HASQI score before the HA has become unstable? We target an HASQI score of about 0.8 and investigate various

[a]Author to whom correspondence should be addressed.

AFC strategies. We compare the performance of different least mean square (LMS) based methods, specifically, the normalized LMS (NLMS),[5] the normalized filtered-X LMS (NFXLMS),[6,7] and the proportionate NLMS (PNLMS).[8,9] Our current findings indicate that an ASG of 12 dB is achievable for an HASQI of 0.8 with a reasonable computational burden.

2. Methods

Figure 1. shows the typical AFC framework. The AFC filter $W(z)$ is a finite impulse response filter placed in parallel with the HA processing $G(z)$ that continuously adjusts its coefficients to emulate the impulse response (IR) of the feedback path $F(z)$. $x(n)$ is the desired input signal and $d(n)$ is the actual input to the microphone, which contains $x(n)$ and the feedback signal $y(n)$ generated by the HA output $s(n)$ passing through $F(z)$. $\hat{y}(n)$ is the estimate of $y(n)$ given by the output of $W(z)$. $e(n) = d(n) - \hat{y}(n)$ is the feedback-compensated signal which, ideally, should be identical to $x(n)$. In practice, however, the AFC is not perfect and therefore $\hat{y}(n) \neq y(n)$, resulting in distortion between $e(n)$ and $x(n)$. The coefficient adaptation for the AFC filter is usually realized by LMS-based algorithms due to computational simplicity and their effectiveness. Finally, $A(z)$ and $H(z)$ are the pre-filter and the band-limited filter present in the NFXLMS approach,[6,7] respectively, as discussed later.

In the following, we will first give a brief introduction of the HASQI. Then the proposed method for estimating ASG is described. Finally, we provide details of the algorithms used in our simulation for verifying the ability of the proposed ASG estimation approach.

2.1 HASQI

The impact of acoustic feedback on perceived speech quality is estimated using the HASQI speech quality metric.[3] The HASQI metric was trained on a large database of subject quality ratings, including nonlinear distortion and frequency response modifications that duplicated the resonance peaks typical of acoustic feedback. The metric was validated on data from a feedback cancellation experiment,[10] and a value of 0.93 was found for the correlation coefficient between the subject ratings and the HASQI quality predictions.[3] In addition, recent papers have shown high degrees of correlation for perceptual metrics used to predict quality ratings for feedback cancellation in HAs.[11,12] However, the idea of using HASQI as an objective metric for optimizing AFC is novel.

HASQI compares the processed HA signal to a reference signal. In this paper, the reference signal is the unmodified computer audio file $x(n)$, and the processed HA signal is the feedback-compensated signal $e(n)$. Both the reference and processed signals are passed through a model of the auditory periphery. The auditory model includes auditory frequency analysis, the dynamic range compression mediated by the outer hair cells, two-tone suppression, and the firing-rate adaptation present in the inner hair cell neural response. The metric compares the time-frequency envelope modulation, temporal fine structure, and long-term spectra between the processed and reference signals to produce the quality prediction. The HASQI model represents a distillation of listener ratings for a large number of linear filtering, noise, and distortion conditions. Since the metric was fit to these responses, the perceptual ratings are built into the predicted quality scores. In addition, HASQI has been validated by several perceptual quality experiments.[3,13–15] HASQI is therefore sensitive to changes in the speech spectrum introduced

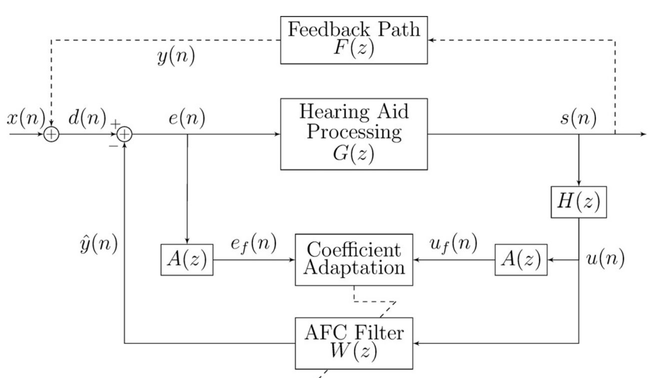

Fig. 1. Block diagram of the AFC framework.

by acoustic feedback, whistling or ringing in the HA, and any nonlinear distortion introduced by the feedback-cancellation processing.

2.2 Proposed approach to ASG estimation using HASQI

For the purpose of estimating the ASG of the AFC algorithm, a uniform gain of the HA processing over all the sub-bands is applied. That is, we use $G(z) = gz^{-d}$, where g represents the gain of the HA and d corresponds to the HA processing delay. The ASG by definition is given as the difference between the MSG of the system with the use of the AFC algorithm and that without the use of AFC (in dB):

$$ASG = MSG_{w/AFC} - MSG_{w/oAFC}. \tag{1}$$

To obtain the ASG estimate, we propose the following procedure.

(i) Define a threshold $T \in (0, 1)$.
(ii) Start from $g = 1$,
  (a) Run the AFC algorithm on a given audio signal $x(n)$ and obtain the feedback-compensated signal $e(n)$.
  (b) Compute the HASQI of $e(n)$ using $x(n)$ as the reference signal. Record the obtained HASQI score.
  (c) If the obtained score $\geq T$,
    Increase g by some small increment, e.g., $\Delta g = 0.1$, and then repeat from (a)
    Else,
    Use the previous g value as the estimate of the MSG. Terminate.
(iii) Perform (ii) for both with AFC and without AFC cases to obtain $MSG_{w/AFC}$ and $MSG_{w/oAFC}$, respectively. Use Eq. (1) (convert into dB first) to obtain the ASG estimate of the AFC algorithm.
(iv) Repeat (ii) and (iii) for multiple audio files. Average over the resulting ASG numbers to obtain the final ASG estimate.

We will also obtain a quality vs gain curve once the above procedure has been done for a particular AFC algorithm with a given audio file. Typically, the quality score will decrease as the gain value increases. Once the score falls below the predefined threshold T, the speech quality is considered unacceptable: We therefore consider the gain at which the system enters the unacceptable state as the MSG of the system.

In our work, a HASQI score of 0.8 was used as the threshold for acceptable/unacceptable states. The HASQI value of 0.8 is consistent with a high quality rating as reported for HA quality evaluations.[13] Because the data are simulation results with no other sources of noise or distortion, the maximum possible HASQI score is 1.0; a value of 0.8 thus represents a measurable degradation in signal quality. Nevertheless, the proposed methodology can still be used for any value of HASQI. For example, a resource constrained HA may target lower speech quality to save power.

The proposed approach was applied to and verified by the ASG estimation for the AFC algorithms described in the following subsection. Note that ASG measured with this approach will be lower than that with the ANSI 3.22 protocol as the later does not consider speech quality prior to entering unusable region.

2.3 LMS-based algorithms for AFC

Here we briefly describe the LMS-based algorithms used in our simulations for verifying the proposed file-based ASG estimation approach. We start with the ordinary LMS, where the band-limited filter and the pre-filter are not present, i.e., $H(z) = A(z) = 1$ in Fig. 1. Let n refer to the sample index and L be the AFC filter length. The LMS update rule for the coefficients of the L-tap AFC filter $\mathbf{w}(n) = [w_0(n), w_1(n), ..., w_{L-1}(n)]^T$ is simply

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu \mathbf{s}(n) e(n), \tag{2}$$

where $\mathbf{s}(n) = [s(n), s(n-1), ..., s(n-L+1)]^T$ and $\mu > 0$ is the step size parameter. In LMS class of algorithms, the step size parameter μ controls the rate at which the AFC parameters are adapted. Larger values of μ result in faster convergence, and smaller values of μ result in higher ASG. Different algorithms have different behaviors to changing feedback paths in real world. In practice, selecting the adaptation parameters is a trade-off based on both technical and commercial reasons. In this work, we focus on ASG and speech quality. Hence, the adaptation parameters were chosen for maximum ASG for a given quality for the AFC algorithms.

NLMS. The ordinary LMS has the disadvantage that it is very sensitive to the input signal level, which easily leads to instability and therefore makes the LMS inapplicable to AFC in practice. The (modified) NLMS[5] is a variant of the LMS algorithm that mitigates this problem by normalizing the step size parameter with the power estimate of the input at time n, which reduces the negative effect brought by large power fluctuations or signal onsets.[2] In the NLMS method, a time-varying step size $\mu(n)$ is employed to improve the convergence rate

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu(n)\mathbf{s}(n)e(n), \tag{3}$$

where

$$\mu(n) = \frac{\mu}{\text{signal power estimate at time } n}. \tag{4}$$

Several methods can be used for estimating the signal power, e.g., using a first order recursive equation.[2,5,7]

NFXLMS. It is well known that the LMS-based AFC in HAs suffers from the bias problem caused by the high correlation between the incoming signal and the feedback signal.[6] As a result, the cancellation performance of the NLMS is still limited for feedback control in HAs. To alleviate this problem the NFXLMS algorithm is applied.[6,7] In the NFXLMS approach, the pre-filter $A(z)$ is utilized to whiten (flatten) the spectra of the input signals to the coefficient adaptation stage in order to reduce their correlation. This pre-filter is usually the inverse of some low pass all-pole filter that models the long term average speech spectrum. Ideally, by decorrelating the signals it will improve the convergence behavior over the NLMS. In addition, the band-limited filter $H(z)$ is also employed to concentrate on the frequency regions where oscillation is likely to occur.[7] Typically, $H(z)$ is a high pass filter and can also be viewed as a very rough approximation of the feedback path in the frequency domain.[6] With the above filters included, the NFXLMS is known to provide better feedback cancellation, convergence speed, and output quality. Now since we have included $A(z)$ and $H(z)$, the NFXLMS update rule is given as

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu(n)\mathbf{u}_f(n)e_f(n), \tag{5}$$

where $\mathbf{u}_f(n) = [u_f(n), u_f(n-1), ..., u_f(n-L+1)]^T$.

NFXLMS with proportionate adaptation. Observing that typical feedback path IRs are sparse (to some degree) as, for example, the one shown in Fig. 2, one might think of taking advantage of this sparseness for AFC improvements. In fact, this can be carried out by the concept of proportionate adaptation, which originated from the PNLMS algorithm.[8] The idea behind proportionate adaptation is to update each filter coefficient independently of the others by assigning to the corresponding step size a weight in proportion to the magnitude of the estimated coefficient. In other words, it redistributes the adaptation gains among all coefficients and emphasizes the large ones in order to speed up their convergence. Recently, attempts have been made to incorporate proportionate adaptation into AFC and improvements have been reported. This results in the proportionate NFXLMS (PNFXLMS) which, in general, can be described by the following update rule:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu(n)\mathbf{P}(n)\mathbf{u}_f(n)e_f(n), \tag{6}$$

where the "proportionate matrix"

$$\mathbf{P}(n) = diag\{p_0(n), p_1(n), ..., p_{L-1}(n)\} \tag{7}$$

is an L-by-L diagonal matrix assigning different weights to the step sizes for different filter taps. In Eq. (7), each $p_l(n)$ is a function of the current AFC filter coefficient

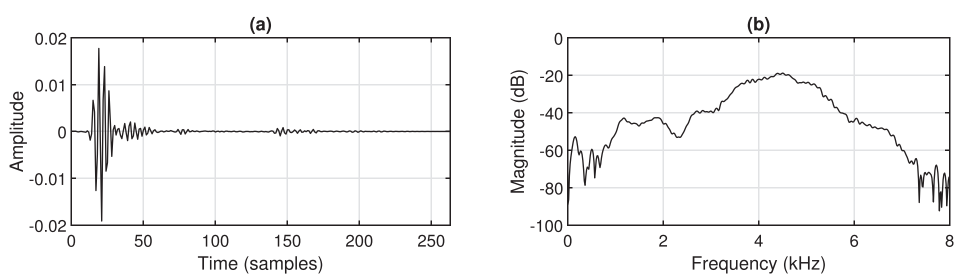

Fig. 2. The (a) impulse response and (b) frequency response of the feedback path used in the simulation.

$w_l(n)$ and is updated every iteration. Different proportionate adaptation algorithms differ in the way updating $p_l(n)$. In this paper, we adopt the approach proposed in Ref. 9.

3. Results

10 audio signals (6 speech signals and 4 music signals) at a sampling rate of 16 kHz with a real-world feedback path IR were used to evaluate the proposed ASG approach. We chose three male and three female speakers from the TIMIT database popular in many speech processing studies. The four music signals (of classical, jazz, choir, and pop genres) were highly tonal signals as tonal content presents challenging conditions for AFC because of high correlation between the input signal and the feedback signal. The feedback path IR was measured without a telephone handset present, truncated and resampled to a length of 263 samples for 16 kHz sampling rate. The IR is shown in Fig. 2, along with its frequency response. All the simulations were conducted using MATLAB with the following settings.

- The HA processing: $G(z) = gz^{-d}$, where $d = 128$ corresponds to an overall system latency of 8 ms and g is the uniform gain (linear scale) of the HA.
- AFC filter length $L = 100$ to cover the significant part of the feedback path IR.
- The pre-filter $A(z)$ and the band-limited filter $H(z)$ are same as used in Ref. 6.
- Step size parameters: $\mu = 10^{-5}$ for NLMS and $\mu = 5 \times 10^{-4}$ for NFXLMS and PNFXLMS.

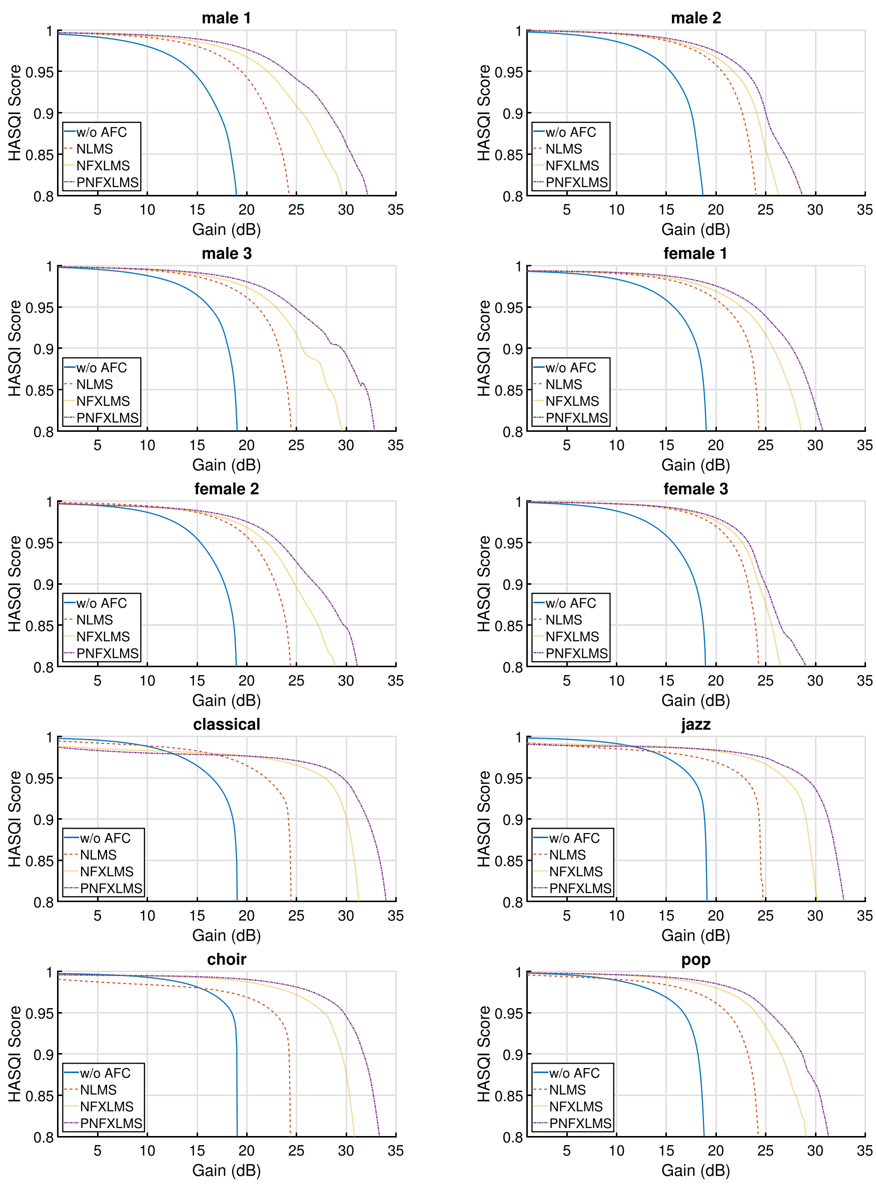

Fig. 3. (Color online) The resulting HASQI curves of the AFC algorithms.

Table 1. Estimated MSG and ASG (in dB) of the AFC algorithms.

Input File	MSG w/o AFC	MSG (NLMS)	ASG (NLMS)	MSG (NFXLMS)	ASG (NFXLMS)	MSG (PNFXLMS)	ASG (PNFXLMS)
male 1	18.89	24.24	**5.35**	29.63	**10.74**	32.15	**13.26**
male 2	18.59	23.97	**5.38**	26.32	**7.73**	28.69	**10.10**
male 3	18.99	24.40	**5.41**	29.63	**10.64**	32.83	**13.84**
female 1	18.99	24.24	**5.25**	28.56	**9.57**	30.73	**11.74**
female 2	18.89	24.35	**5.46**	28.85	**9.96**	31.10	**12.21**
female 3	18.89	24.24	**5.35**	26.49	**7.60**	29.04	**10.15**
classical	18.99	24.40	**5.41**	31.25	**12.26**	34.01	**15.02**
jazz	19.08	24.71	**5.63**	30.18	**11.10**	32.85	**13.77**
choir	18.99	24.35	**5.36**	30.81	**11.82**	33.33	**14.34**
pop	18.79	24.19	**5.40**	29.04	**10.25**	31.29	**12.50**
Average	18.91	24.31	**5.40**	29.08	**10.17**	31.60	**12.69**

For each test file, we varied the gain value g of the HA processing with $\Delta g = 0.1$ increment and measured the corresponding HASQI score between the reference signal $x(n)$ and the processed signal $e(n)$ for each g, under four different situations: (i) without AFC, (ii) with NLMS, (iii) with NFXLMS, and (iv) with PNFXLMS. The adaptive filter $\mathbf{w}(n)$ was initialized as a proper estimate of the feedback path, which was obtained by averaging several feedback path IRs measured under different scenarios (e.g., without a telephone handset present and with the handset positioned in different distances from the ear). The resulting HASQI curves are plotted in Fig. 3 for all the test files. Note that in Fig. 3 we have converted the gain values into dB scale. For each HASQI curve in each plot, we can identify the MSG as the gain value before the curve falls below 0.8. The obtained MSG numbers and computed ASG numbers are tabulated in Table 1 for all the test files.

We also measured the average runtimes for the AFC processing stage to assess the computational burden for the above algorithms. We normalized the runtimes with respect to the numbers of the NLMS. For the content used in Table 1, NFXLMS requires additional 8% (standard deviation 1.68%) and PNFXLMS 11% (standard deviation 1.69%) on a laptop compared to the NLMS. It can be concluded that ASG gains of 4 and 7 dB for NFXLMS and PNFXLMS (at HASQI = 0.8) can be achieved at a computational burden of additional 8% and 11%, respectively, over the NLMS. Note that the computational burden can vary significantly in embedded implementations, depending on the nature of implementation optimizations.

4. Conclusions and discussion

In this paper, a method that takes advantage of HASQI to measure the ASG of feedback cancellation algorithms for HAs is proposed and investigated. HASQI provides an objective metric that can capture HA performance prior to entering unusable region as opposed to ANSI 3.22 protocol that does not consider speech quality. We show that the HASQI score is an effective metric for evaluating the audio quality of the feedback-compensated signal by testing three different LMS-based AFC algorithms: the NLMS, the NFXLMS, and the PNFXLMS, on several speech and music signals. In the NFXLMS, the pre-filter and band-limited filter are expected to improve feedback cancellation performance over the NLMS due to decorrelation of the signals and concentration on oscillation frequency regions. Furthermore, with the proportionate adaptation included for exploiting sparsity in the feedback channel, the PNFXLMS is expected to provide even better feedback cancellation ability. The proposed ASG estimation approach well reflects these improvements. To conclude, this file-based ASG estimation approach that utilizes HASQI is capable of reflecting the feedback cancellation ability in the stable region of the HA operation and hence can serve as a good way to compare different AFC algorithms.

Acknowledgments

This work was supported by NIH/NIDCD Grants Nos. 1R21DC015046 and 1R01DC015436.

References and links

[1]T. van Waterschoot and M. Moonen, "Fifty years of acoustic feedback control: State of the art and future challenges," Proc. IEEE **99**(2), 288–327 (2011).

[2]A. Spriet, S. Doclo, M. Moonen, and J. Wouters, "Feedback control in hearing aids," in *Springer Handbook of Speech Processing* (Springer, Berlin, 2008), pp. 979–1000.
[3]J. M. Kates and K. H. Arehart, "The hearing-aid speech quality index (HASQI) version 2," J. Audio Eng. Soc. **62**(3), 99–117 (2014).
[4]J. M. Kates, "Constrained adaptation for feedback cancellation in hearing aids," J. Acoust. Soc. Am. **106**(2), 1010–1019 (1999).
[5]J. E. Greenberg, "Modified LMS algorithms for speech processing with an adaptive noise canceller," IEEE Trans. Speech Audio Process. **6**(4), 338–351 (1998).
[6]J. Hellgren, "Analysis of feedback cancellation in hearing aids with Filtered-X LMS and the direct method of closed loop identification," IEEE Trans. Speech Audio Process. **10**(2), 119–131 (2002).
[7]H.-F. Chi, S. X. Gao, S. D. Soli, and A. Alwan, "Band-limited feedback cancellation with a modified filtered-X LMS algorithm for hearing aids," Speech Commun. **39**(1-2), 147–161 (2003).
[8]D. L. Duttweiler, "Proportionate normalized least-mean-squares adaptation in echo cancelers," IEEE Trans. Speech Audio Process. **8**(5), 508–518 (2000).
[9]C. Paleologu, J. Benesty, and S. Ciochină, "An improved proportionate NLMS algorithm based on the l_0 norm," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)* (2010), pp. 309–312.
[10]J. Bondy, A. Dittberner, M. Coughlin, and B. Whitmer, "Assessing sound quality of feedback algorithms with auditory models," in *International Symposium on Auditory and Audiological Research*, Helsingør, Denmark (August 29–31, 2007).
[11]A. J. Manders, D. M. Simpson, and S. L. Bell, "Objective prediction of the sound quality of music processed by an adaptive feedback canceller," IEEE Trans. Audio Speech Lang. Process. **20**(6), 1734–1745 (2012).
[12]M. Guo, S. H. Jensen, and J. Jensen, "Evaluation of state-of-the-art acoustic feedback cancellation systems for hearing aids," J. Audio Eng. Soc. **61**(3), 125–137 (2013), see http://www.aes.org/e-lib/browse.cfm?elib=16673.
[13]D. Suelzle, V. Parsa, and T. H. Falk, "On a reference-free speech quality estimator for hearing aids," J. Acoust. Soc. Am. **133**(5), EL412–EL418 (2013).
[14]A. A. Kressner, D. V. Anderson, and C. J. Rozell, "Evaluating the generalization of the hearing aid speech quality index (HASQI)," IEEE Trans. Audio Speech Lang. Process. **21**(2), 407–415 (2013).
[15]P. Kendrick, I. R. Jackson, F. F. Li, T. J. Cox, and B. M. Fazenda, "Perceived audio quality of sounds degraded by non-linear distortions and single-ended assessment using HASQI," J. Audio Eng. Soc. **63**(9), 698–712 (2015).