Evaluation of State-of-the-Art Acoustic Feedback Cancellation Systems for Hearing Aids

MENG GUO1,2, SØREN HOLDT JENSEN2 AND JESPER JENSEN1

1Oticon A/S, Kongebakken 9, DK-2765 Smørum, Denmark
2Department of Electronic Systems, Aalborg University, Niels Jernes Vej 12, DK-9220 Aalborg, Denmark

In this work we evaluate four state-of-the-art acoustic feedback cancellation systems for hearing aid applications. We show that significant improvements in cancellation performance can be made over traditional systems by allowing small alterations in the loudspeaker signal and a computational complexity increase by a factor of 2 – 3. The evaluation is based on a listening test and objective assessments of simulation results.

0 INTRODUCTION

Acoustic feedback occurs when the output signal of an audio device returns to its microphone and thereby forms an acoustic feedback loop. The typical consequences of acoustic feedback are sound quality degradation and, in the worst-case, howling. Acoustic feedback occurs typically in a sound reinforcement system such as public address systems and hearing aids; especially, it is very likely to occur in a hearing aid, due to the closely placed microphones and loudspeaker, typically only a few millimeters to a few centimeters apart depending on the hearing aid style. In this work we focus on hearing aid systems.

Acoustic feedback cancellation (AFC) using adaptive filter techniques in a system identification configuration [1] has become the state-of-the-art method for reducing the effect of acoustic feedback [2]. Fig. 1 illustrates a simple hearing aid system with an AFC system, where an adaptive filter \( \hat{h}(n) \) models and cancels the acoustic feedback path \( h(n) \) from the hearing aid loudspeaker to the microphone. The hearing aid processing is represented by the generally time-varying forward path impulse response \( f(n) \), and the microphone signal \( y(n) \) consists of the desired incoming signal \( x(n) \) and the undesired but unavoidable feedback signal \( v(n) \), whereas the loudspeaker signal is denoted by \( u(n) \).

In a traditional fullband (FB) AFC system [2], the cancellation is carried out by updating the adaptive filter \( \hat{h}(n) \) using normalized least mean square (NLMS), affine projection (AP), or recursive least squares (RLS) algorithms.

1 In hearing aid terminology, the loudspeaker is generally referred to as the receiver. However, since several of the algorithms under study find applications also outside the hearing aid area, we refer to it as the loudspeaker in this work.
that are not always valid in practice (e.g., the PEM based techniques), or they introduce modifications to the loudspeaker signal \( u(n) \) to decorrelate it from \( x(n) \), which might cause sound quality degradations, as, e.g., the frequency shifting and probe noise based approaches [18]. However, allowing a minor sound quality degradation from the decorrelation process can very often improve AFC performance greatly, and it is thereby possible to obtain a much better overall sound quality in \( u(n) \). Furthermore, for practical applications such as hearing aids, another important issue is the computational complexity. Hence, the trade-off in practice is often between getting a better AFC performance by obtaining an unbiased estimation of \( \hat{h}(n) \), the degree of introduced sound quality degradations for obtaining this unbiased estimation, and computational complexity.

Some recent studies have compared the performance of different AFC systems. A comparison of different sound reinforcement AFC systems was presented in [18], where the cancellation performance and sound quality degradation are both evaluated objectively. In [19], a comparison based on physical measurements of different commercial hearing aids is reported. In this study, however, it is not clear which AFC system is used in each individual commercial hearing aid, and these AFC systems might not be state-of-the-art. Hence, no reported work exists for comparing different state-of-the-art AFC algorithms/systems for hearing aids, in terms of AFC performance, sound quality degradation, and computational complexity. Moreover, the recent introduction of a novel probe noise based AFC system [17] makes such a comparison even more appealing.

Therefore, in this work, we compare four different state-of-the-art AFC systems to a traditional fullband AFC system in terms of their abilities to cancel acoustic feedback, the sound quality degradations they might introduce to decorrelate the loudspeaker signal from the incoming signal, and their computational complexity. This comparison is performed using simulation experiments with realistic hearing aid setups and objective performance measures. To ensure a high sound quality in \( u(n) \), the distortions (if any) introduced in \( u(n) \) for decorrelation must be controlled; this is carried out by choosing appropriate parameters in different systems, and we evaluate the sound quality using both objective quality measures and a subjective listening test.

The rest of this paper is organized as follows. In Section 1 we provide an overview of all considered AFC systems.

### 1 OVERVIEW OF DIFFERENT AFC SYSTEMS

In this work we consider two different structures for adaptive filter estimation: a fullband NLMS algorithm and a subband (SB) NLMS algorithm. Other adaptive algorithms such as AP and RLS can be used for the adaptive estimation of \( \hat{h}(n) \), but the NLMS algorithm is often chosen in hearing aid applications for its simplicity. A subband system is more complex in structure than the fullband system, but it has more freedom in its configuration since one step size parameter per subband is available for the adaptive algorithm compared to a single step size parameter for all frequencies in the fullband system. In addition, due to the subband processing, each subband spectrum has generally a smaller spectral dynamic range than the fullband spectrum for a nonwhite signal, and an increased convergence rate of the adaptive estimation is thereby possible [20]. Furthermore, in contrast to the fullband system, a subband system is often less sensitive to the biased estimation problem, which would only affect a few subbands in the frequency regions with strong signal correlations between \( u(n) \) and \( x(n) \), while a correct estimation would still be possible in the remaining frequency regions.

Each adaptive filter structure can be further combined with different decorrelation methods to form a complete AFC system. We focus on the prediction error method because it has been shown to have superior performance compared to many other decorrelation methods [18], and it does not introduce sound degradations in \( u(n) \). Furthermore, we focus on the frequency shifting decorrelation method for its performance [18] and simplicity [21]. Finally, we consider the recently proposed probe noise approach in combination with probe noise enhancement [17].

<table>
<thead>
<tr>
<th></th>
<th>No Decor.</th>
<th>PEM</th>
<th>FS</th>
<th>PN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fullband</td>
<td>✔</td>
<td>✔</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Subband</td>
<td>✔</td>
<td>–</td>
<td>✔</td>
<td>✔</td>
</tr>
</tbody>
</table>

In Section 2 we evaluate the sound quality for two systems that introduce additional sound degradations for decorrelation. In Section 3 we perform simulation experiments, using audio signals, to compare cancellation performance for all systems. Section 4 outlines the computational complexity of the systems. Finally, we conclude this work in Section 5.
evaluation of a simple alternative to the subband system with probe noise for decorrelation as described in [17]. Using computer simulations with floating point arithmetic, we evaluate these five systems in a hearing aid application with realistic parameter settings.

In the following, we describe each of the five AFC systems. Table 2 provides descriptions and abbreviations for the five systems. Fig. 2 shows a compact block diagram for all five AFC systems. For each system I-V, the blocks A-D in the block diagram should be accordingly substituted by their replacement blocks given in the dashed boxes A-D.

### 1.1 System I: F-AFC System

In the F-AFC system [2], the adaptive filter \( \hat{h}(n) \) is estimated to minimize the mean square error \( E[\hat{e}^2(n)] \). An NLMS algorithm is used to estimate \( \hat{h}(n) \). As we will demonstrate in Section 3, one of the major limitations in this system is the biased estimation of \( \hat{h}(n) \) whenever \( u(n) \) is correlated with \( x(n) \).

### 1.2 System II: PEM-AFC System

The PEM-AFC systems [11,12] assume that the incoming signal \( x(n) \) can be modeled as a white noise sequence filtered through an autoregressive model

\[
1/ \sum_{k=0}^{L-1} p_k(n)z^{-k},
\]

where \( z^{-1} \) is the unit delay operator. Let \( p(n) = [p_0(n), p_1(n), \ldots, p_{L-1}(n)]^T \), the prefilter \( \hat{p}(n) = [\hat{p}_0(n), \hat{p}_1(n), \ldots, \hat{p}_{L-1}(n)]^T \) and the cancellation filter \( \hat{h}(n) \) are then jointly estimated to minimize the mean square error \( E[\hat{e}^2(n)] \), where \( \hat{e}_p(n) = \sum_{k=0}^{L-1} \hat{p}_k(n)e(n-k) \) is the filtered signal of \( e(n) \).

For an ideal estimation, \( \hat{p}(n) = p(n) \), the filtered error signal \( \hat{e}_p(n) \) only contains the filtered signal \( v_p(n) \) of \( v(n) \) and samples of the white noise excitation sequence, which are uncorrelated with the loudspeaker signal \( u(n) \) and the filtered signal \( u_p(n) \) of \( u(n) \), respectively. In this way, the PEM-AFC ideally provides an unbiased estimation of \( \hat{h}(n) \) for autoregressive incoming signals \( x(n) \).

In different versions of the PEM-AFC system, \( x(n) \) is modeled in different ways, and various schemes to estimate \( \hat{p}(n) \) and \( \hat{h}(n) \) have been proposed. For more details on the PEM-AFC system, we refer to [4,11,12].

### 1.3 System III: S-AFC System

The S-AFC system facilitates adaptive filter estimation in filter bank subbands. Several variants of subband and frequency domain adaptive filter processing are possible, e.g., [22,23]. We employ the delayless subband structure proposed in [24] and refined in [25]. The greatest advantage by using this structure in contrast to other subband adaptive filter structures, e.g., [22], is that no additional delay is introduced in the signal path from \( y(n) \) to \( u(n) \). In particular, uniform filter banks are used to divide the fullband error signal \( e(n) \) and loudspeaker signal \( u(n) \) to subband signals \( E(m,k) \) and \( U(m,k) \) with the subband frequency index \( m \) and time index \( k \), respectively. The adaptive filter estimation of the frequency response \( H(m,k) \) is then performed in the subband domain using a low-order NLMS algorithm based

![Fig. 2. An overview of the AFC systems I-V. For simplicity, we only consider a hearing aid system with a single microphone. For each system I-V, the blocks A-D in the block diagram should be accordingly substituted by their replacement blocks given in the dashed boxes A-D.](image-url)
on $E(m,k)$ and $U(m,k)$. The fullband adaptive filter estimate $\hat{h}(n)$ is obtained by the inverse discrete Fourier transform of the frequency response estimate $\hat{H}(m,k)$.

The subband structure makes it easy to use different parameter settings across frequency regions, e.g., the subband step size parameter. For hearing aid applications, AFC systems are often only necessary for frequencies higher than above approximately 1 kHz [11]. On the other hand, the signal correlation between $u(n)$ and $x(n)$ is often most dominant at low frequency regions, e.g., for speech signals. Therefore, having AFC systems running in the lowest frequencies might introduce artifacts due to the biased estimation problem rather than solving a small acoustic feedback problem. In this subband structure, AFC systems at lower frequencies can easily be turned off by setting the corresponding subband step sizes to zero.

1.4 System IV: FS-AFC System

The FS-AFC system utilizes the same subband AFC system as the S-AFC system. In addition, a frequency shifting [21] is applied before the hearing aid processing unit $f(n)$. It is known that frequency shifting decorrelates the loudspeaker signal $u(n)$ from the incoming signal $x(n)$ and can thereby reduce the biased estimation problem [14,15,18].

1.5 System V: PN-AFC System

The PN-AFC system also makes use of the subband AFC system, and it is described in detail in [17]. Instead of modifying the loudspeaker signal $u(n)$ via frequency shifting, this system obtains an unbiased estimation of adaptive filter $\hat{h}(n)$ by adding a generally non-stationary probe noise signal $w(n)$, which is uncorrelated with $x(n)$, to the loudspeaker signal $u(n)$. The probe noise signal $w(n)$ is generated using a masking model to ensure that $w(n)$ is inaudible in the presence of the loudspeaker signal $u(n)$.

Furthermore, enhancement filters $\hat{a}(n)$ are used for filtering the probe noise signal $w(n)$ and the error signal $e(n)$ before they enter the adaptive estimation algorithm. The enhancement filters have a specific structure that allows them to improve the probe noise to disturbing signal power ratio and then improve the estimation of $\hat{h}(n)$. We refer to [17] for more details.

2 SOUND QUALITY EVALUATION OF DECORRELATION METHODS

Among the five AFC systems the FS-AFC and PN-AFC systems might introduce additional audible artifacts because they modify the loudspeaker signal $u(n)$ to decorrelate it from $x(n)$. More specifically, for the FS-AFC system, the loudspeaker signal $u(n)$ is frequency shifted, whereas the PN-AFC system adds a probe noise signal $w(n)$ to the loudspeaker signal $u(n)$. These modifications of the loudspeaker signal $u(n)$ might be perceived as a sound quality degradation. Depending on the exact system parameter settings, frequency shifting might introduce a vibrato-like sound degradation for both speech and music signals, whereas probe noise might add an audible hiss to the original signal. Therefore, it is important to understand how these modifications of the loudspeaker signal $u(n)$ affect the sound quality. Ideally, only minor and nonirritating audible artifacts should be allowed to preserve sound quality. Furthermore, the degradations should ideally be at similar levels for making the AFC performance comparison in Section 3 more straightforward. To achieve these, we perform a sound quality evaluation.

2.1 Evaluation Method

We carry out a listening test to evaluate the sound quality degradation due to the decorrelation in terms of frequency shifting and probe noise injection in the FS-AFC and PN-AFC systems, respectively. We initially adjust system parameters in both decorrelation methods so that they are effective for decorrelation, and the sound quality degradations are in the range of intermediate to small (verified by informal listening tests). Moreover, we want to evaluate the sound quality for both speech and music signals. To achieve this, we use the “MULTi Stimulus test with Hidden Reference and Anchor (MUSHRA)” specified in [26] with one reference signal, five test signals including a hidden reference, a hidden anchor, two test signals processed with frequency shifting variants, and one test signal processed with probe noise injection. In the MUSHRA, the test signals are compared to each other and to the reference and the anchor signals. A different approach to this is the absolute category rating [27], where test subjects should give a quality rating based on a single test signal. However, both listening test approaches are somewhat related to each other, as discussed in [28].

The sound quality is evaluated using normal hearing test subjects. Our underlying hypothesis is that if the sound quality degradation is not significant nor annoying for normal hearing subjects, then it would be acceptable for hearing impairment subjects too. In this way, we expect the results of normal hearing test subjects to provide a sound quality acceptance lower bound. The studies in [29,30] support this assumption.

The listening test was conducted during a period of three days in a quiet room, and a total of 16 normal-hearing test subjects participated in the test. All test subjects were male between 24 and 57 years old, and they were experienced in judging/assessing sound distortions. The test signals were presented diotically via a headphone (Sennheiser EH2270, frequency response: 12 – 22000 Hz, total harmonic distortion: <0.2%) connected to a computer, which ran the MUSHRA interface, and the sound level was adjusted individually by the subjects. The rating of each test signal was recorded by the computer.

2.2 Processing of Test Signals

We focus on the artifacts of the frequency shifting and probe noise injection, without any interaction with AFC systems; it makes the interpretation of this sound quality evaluation results more straightforward. Moving back to Fig. 2, it corresponds to removing the feedback path $h(n)$ and its estimate $\hat{h}(n)$, i.e., $h(n) = 0$ and $\hat{h}(n) = 0$. 

The reference signal in the MUSHRA is created by passing an unprocessed high quality sound signal through the hearing aid processing with only an amplification and delay. Test signals are created by passing the same high quality sound signal through the hearing aid processing with the same amplification and delay and an additional frequency shifting or probe noise injection; in this way, the hearing aid loudspeaker signals \( u(n) \) are presented to test subjects in this listening test.\(^2\)\(^\text{1}\) The anchor signal is created in a similar way as the test signals but with both frequency shifting and probe noise injection. Table 3 shows the five sound signals chosen for the MUSHRA. Details on the processing conditions are given in Table 4.

To ensure that the chosen parameters for frequency shifting and probe noise injection given in Table 4 do not introduce significant sound quality degradations, we initially verified these parameter choices via informal listening tests and using the objective evaluation measures PESQ [32] and PEAQ [33] for, respectively, speech and music signals (although these objective evaluation measures were not specifically designed for evaluating the artifacts of either frequency shifting nor probe noise, an alternative to PESQ and PEAQ measures could be the hearing aid speech quality index [34]). Fig. 3 shows the PESQ/PEAQ prediction results.

The PESQ/PEAQ predictions were in line with the observations made via the initial informal listening, and they indicate that the parameter choices actually provide non-significant sound quality degradations. The minor sound quality degradations from both methods are at a similar level, which makes the AFC performance evaluation results in Section 3 more directly comparable.

\(^2\) We considered to present instead a mixture of the loudspeaker signal \( u(n) \) and the incoming signal \( x(n) \) because this would be closer to reality for some hearing aid styles [31]. However, with the mixed signal an undesired comb filter effect is probable to occur [31]. The comb filter effect makes the listening test more complicated, since test subjects have to assess and judge one type of artifact (comb filter effect) found in the reference signal from other types of artifacts (comb filter effect + frequency shifting or probe noise) found in the test signals. A potential risk of doing such a test is that test subjects would non-deliberately pay more attention to the comb filter effect instead of artifacts from frequency shifting and probe noise. Therefore, to avoid this, we decided to present the loudspeaker signals \( u(n) \) in the MUSHRA.

### Table 3. Sound signals used for the MUSHRA.

<table>
<thead>
<tr>
<th>Sound Signal</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speech 1</td>
<td>Danish Female Reading</td>
</tr>
<tr>
<td>Speech 2</td>
<td>English Male Reading</td>
</tr>
<tr>
<td>Music 1</td>
<td>Piano + Female Singing</td>
</tr>
<tr>
<td>Music 2</td>
<td>Female Singing</td>
</tr>
<tr>
<td>Music 3</td>
<td>Singing Female Voice</td>
</tr>
</tbody>
</table>

### 2.3 Training and Test Procedure

In subjective tests including the MUSHRA, training of test subjects is important to obtaining reliable results. Preliminary tests and experiences gathered through the development of the FS-AFC system show that the training for this listening test is difficult.

Specifically, we noticed that the training period in the frequency shifting case might be as long as days or weeks of listening, before the artifacts become clearly noticeable for even experienced test subjects. However, once these artifacts become noticeable, they appear to be very annoying for some test subjects. On the other hand, we did not notice that test subjects became more annoyed with probe noise injection over time. Detailed studies of this time course could be a topic for future work.

For practical reasons in this work, we were not able to conduct a preparation over weeks for a large group of test subjects. Therefore, we decided to have two training sessions prior to the actual test session, where the artifacts of frequency shifting and probe noise are exaggerated in the first training session in the hope that this would shorten the training period. The three test signals are processed by frequency shifting of 20 Hz and probe noise threshold 1.5 · \( M(m, k) \) in contrast to the nominal 10 Hz and \( M(m, k) \) as given in Table 4, which are used for the second training session and the test session. In this way, the test consists of three sessions: two training sessions and a test session.

### 2.4 Listening Test

We instructed test subjects to find the hidden reference and anchor for each test trial and rate them with the quality scores 100 and 0, respectively. The remaining test signals should be rated with respect to the reference and anchor. We did not inform test subjects which kind of artifacts or the processing algorithms they were listening to. For most subjects, the entire listening test took about 20 to 40 minutes including the initial instruction, two training sessions, the test session, and a short debriefing at the end of the test.
2.5 Test Results

Only the results from the test session are used for determining test statistics. To verify the test-retest reliability for the obtained data, we made a simple correlation test between all quality scores given by test subjects in the second training session and the final test session. The computed Pearson correlation coefficient $r = 0.76$ shows that the obtained data are relatively reliable. The statistics determined from all subjects are shown in Fig. 4. Since all test subjects were able to correctly find the hidden reference and anchor, the results of these are not presented. Seven test subjects were already familiar with both type of artifacts before the listening test and can therefore be considered as expert listeners for this test. Fig. 5 shows the statistics of test results from this expert panel.

### Table 4. Processing of test signals for each test sound. All test signals are processed with identical amplification and delay.

<table>
<thead>
<tr>
<th>Processing</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference</td>
<td>Amplification + Delay.</td>
</tr>
<tr>
<td>FS-500</td>
<td>Amplification + Delay + Frequency shifting of 10 Hz above 500 Hz.</td>
</tr>
<tr>
<td>FS-1000</td>
<td>Amplification + Delay + Frequency shifting of 10 Hz above 1000 Hz.</td>
</tr>
<tr>
<td>PN-500</td>
<td>Amplification + Delay + Probe noise above 500 Hz with noise level computed based on a spectral masking threshold $M(m, k)$ for the nth frequency at time index $k$, see [17] for more details.</td>
</tr>
<tr>
<td>Anchor</td>
<td>Amplification + Delay + Frequency shifting of 25 Hz above 500 Hz and probe noise injection above 500 Hz with noise level $2 \cdot M(m, k)$.</td>
</tr>
</tbody>
</table>

2.6 Discussions

From Figs. 4 and 5 we conclude that the frequency shifting of 10 Hz with a cut-off frequency at 1000 Hz (FS-1000) and probe noise with a cut-off frequency at 500 Hz (PN-500) introduce relatively small degradations at similar levels, especially for speech signals. Moreover, we conclude that the frequency shifting with a cut-off frequency at 500 Hz (FS-500) generally produces signals with sound quality in the range of poor to fair. We carried out a two-way repeated measures analysis of variance (rANOVA) on the test results for all test subjects, with sound signal and processing condition as the two independent variables, since each test subject provided a sound quality score for each permutation of these two variables. First we considered the hypothesis that the sound quality scores from the processing conditions FS-1000 and PN-500 were equal. The rANOVA showed a p-value of $p \approx 0.64$, and we accept this hypothesis. Furthermore, we did a second rANOVA by pooling the sound quality scores from FS-1000 and PN-500 processing conditions together and tested them against the scores from the FS-500 processing. The obtained p-value of $p < 0.00003$ indicates that FS-500 provides significantly different scores than FS-1000 and PN-500. Recalling that our goal of this experiment was to ensure degradations at similar levels for both frequency shifting and probe noise processing, and only minor and nonirritating audible artifacts should be allowed to preserve sound quality, we therefore exclude the FS-500 processing from the following discussion.

The PN-500 processing seems to have a higher degradation level for speech signals compared to the FS-1000. Nevertheless, the sound quality of speech signals is still perceived as almost excellent for the PN-500. This trend is just the opposite for music signals, where the PN-500 seems to provide a slightly better sound quality than the FS-1000. However, it should be noted that the PN-500 has a decorrelation effect in the frequency region of 500 – 1000 Hz, which is not the case for the FS-1000, and better AFC performance can thereby be expected for the PN-500.

Interestingly, the expert test subjects rated the test signals differently than the general test subjects. By comparing Figs. 4 and 5, it is clear that the expert subjects generally rated the frequency shifting processed test signals (FS-1000) lower, and the probe noise processed test signals (PN-500) higher compared to the general test subjects, although it is not statistically significant due to the relatively small sample size in the test. This can be explained by the hypothesis that a long training period is necessary for test subjects to sufficiently learn to rate test signals in a proper way.

...
subjects to be aware of the frequency shifting artifact, and they tend to dislike it once noticing it. Fig. 5 suggests that the sound quality degradation with FS-1000 is somewhat higher than with PN-500, especially for music signals, and that only two training sessions are not adequate for general test subjects. Moreover, these results shown in Fig. 5 can also be considered as upper-bounds for the sound quality test subjects. Moreover, these results shown in Fig. 5 can also be considered as upper-bounds for the sound quality test subjects. Furthermore, test subjects typically used words such as “vibration,” “ringing sounds,” “modulation,” “metallic,” and “additional tones” to describe the frequency shifting artifacts. Whereas for the probe noise artifacts, the most common describing words are “noisy” and “hiss.” Test subjects also suggested that the level of sound quality degradation depends on test signals, and speech signals generally have higher quality. Most test subjects reported that they get more annoyed by the modulation artifacts from the frequency shifting because they do not “fit into” the sound, whereas the noise from the probe noise processing is better “hidden” in the original signals. Finally, a comparison of Figs. 3 and 5 shows that there is a similarity between the sound quality prediction using PESQ/PEAQ and the subjective evaluated sound quality using the MUSHRA with expert subjects. Similar trends are found in both the PESQ/PEAQ measures and the MUSHRA results, e.g., that the speech signals have generally higher sound quality scores than music signals, and the PN-500 provides a similar if not better overall sound quality compared to the FS-1000 processing. The absolute values of sound quality scores might not be directly comparable between these two evaluations since the sound quality scores in the MUSHRA depend on the anchor signal; increasing the quality of the anchor signal would somewhat lower the sound quality perception for the test signals and vice versa. However, it is still clear that the PESQ/PEAQ measures seem to provide an accurate relative sound quality prediction. Furthermore, the Spearman’s rank correlation coefficient between these measures is determined to be $\rho = 0.51$, which somewhat supports the suggested similarity. In addition to the sound quality consideration, another important concern of these two decorrelation methods regards the preservation of the localization cues for binocular reproductive systems upon modifications of loudspeaker signals. This is a topic for future work.

3 AFC PERFORMANCE EVALUATION

In the previous section evaluations are carried out to ensure that the signal distortions due to decorrelation are insignificant and at similar levels for the FS-AFC and PN-AFC systems. In this section we conduct simulation experiments for comparing the ability to cancel feedback in all five AFC systems.

3.1 Objective Performance Measures

To quantify objectively the cancellation performance, we use two performance measures to be described in the following. The difference between the feedback path impulse response $h(n)$ and its estimate $\hat{h}(n)$ is directly linked to system stability via the open-loop transfer function $\Theta(\omega, n)$. For the general system (ignoring the frequency shifting) shown in Fig. 2, $\Theta(\omega, n)$ is expressed by

$$\Theta(\omega, n) = F(\omega, n) \left( H(\omega, n) - \hat{H}(\omega, n) \right),$$  

where $F(\omega, n)$ is the frequency response of the hearing aid forward path impulse response $f(n)$, $H(\omega, n)$ and $\hat{H}(\omega, n)$ are frequency responses of $h(n)$ and $\hat{h}(n)$, respectively. The open-loop transfer function $\Theta(\omega, n)$ determines system stability according to the Nyquist stability criterion [35] (see, e.g., [6, 36, 4, 2, 37] for application examples), which states that a linear and time-invariant closed-loop system becomes unstable whenever the following two criteria are both fulfilled:

1. $|\Theta(\omega, n)| \geq 1$;  
2. $\angle \Theta(\omega, n) = l \pi$, $l = \mathbb{Z}$.

Equation (2) forms the basis for different distance measures between $h(n)$ and $\hat{h}(n)$ [1]. A direct measure of the convergence rate, tracking error, and the steady-state error of the adaptive filter $\hat{h}(n)$ is the frequency domain coefficient misalignment measure $\varepsilon(\omega, n)$ defined as

$$\varepsilon(\omega, n) = |H(\omega, n) - \hat{H}(\omega, n)|.$$  

Focusing on the most critical value of $\varepsilon(\omega, n)$ in Eq. (4) over frequencies, we define a simplified measure referred to as the maximum coefficient misalignment (MCM) $\varepsilon_F(n)$; it is determined by the maximum of $\varepsilon(\omega, n)$ across a frequency region denoted by $\mathcal{F}$, as

$$\varepsilon_F(n) \text{ [dB]} = 20 \log_{10} \max_{\omega \in \mathcal{F}} \varepsilon(\omega, n).$$  

Furthermore, assuming that Eq. (3) is fulfilled for all frequencies, a conservative maximum fullband ($\omega \in [0, \pi]$) gain to ensure system stability referred to as the maximum stable gain (MSG) $M(n)$ is determined as

$$M(n) \text{ [dB]} = -20 \log_{10} \max_{\omega} \varepsilon(\omega, n).$$  

More details of $M(n)$ can be found in [38], and for descriptions of physical measurements of $M(n)$ we refer to [6].

In this work, we use $\varepsilon_F(n)$ and $M(n)$ for the objective performance evaluation.

3.2 Test Setups and Signals

3.2.1 Acoustic Feedback Path

The acoustic feedback paths $h(n)$ used in the simulation experiments are obtained from measurements on hearing aids. During simulations, the feedback path $h(n)$ remains fixed for most of the time. However, halfway through each simulation run, a feedback path change is simulated by a complete and momentary change of $h(n)$. This change models a situation, where a hearing aid user places a telephone close to the ear and thereby the hearing aid, and it is known to be a very challenging situation for hearing aid AFC systems [39].
With Telephone

The open-loop transfer function $|\Theta(\omega, n)|$ is used to provide a fullband, time-varying, amplification and delay for all systems. A single-channel 132 J. Audio Eng. Soc., Vol. 61, No. 3, 2013 March

3.2.4 AFC Systems

3.2.3 Test Signals

3.2.2 Forward Path

For simplicity, the forward path $f(n)$ only consists of an amplification and delay for all systems. A single-channel compressor [31] is used to provide a fullband, time-varying, and signal dependent amplification. The maximum amplification of the compressor is limited so that the magnitude of the open-loop transfer function $|\Theta(\omega, n)|$ is $-1$ dB without acoustic feedback cancellation, i.e., when $\hat{h}(n) = 0$.

3.2.3 Test Signals

We use speech and music signals as test signals $x(n)$ in the simulation experiments. The duration of each test signal is 60 s, and the sampling rate is 20 kHz.

In this work, we choose speech signals spoken by both male and female speakers and in different languages. In contrast to speech, for which the autocorrelation time is often several tens of ms, music signals might have correlation time of several hundreds or even thousands of ms, and it could easily cause significantly biased estimation of $\hat{h}(n)$. Therefore, we choose some music signals with clear sustaining tonal components for relatively long time periods. With these test signals and the already introduced acoustic feedback path change, we determine cancellation performance of different AFC systems in a very demanding but realistic situation. Table 5 summarizes the test signals used in the simulations.

3.2.4 AFC Systems

For all AFC systems, the length of $\hat{h}(n)$ is 64 at a sampling rate of 20 kHz. Furthermore, we choose the parameters of the various systems so that a similar and relatively fast convergence is achieved for all systems for speech signals.

For the F-AFC system, an NLMS algorithm with a step size of $2^{-9}$ and a regularization parameter of $2^{-12}$ is used to estimate $\hat{h}(n)$.

For the PEM-AFC system, we use the Levinson-Durbin recursion [40] to compute the prefilter $\hat{p}(n)$ of order 20 based on the error signal $e(n)$, whereas we use the same NLMS algorithm and parameter settings for the estimation of $\hat{h}(n)$ as the F-AFC system.

For the S-AFC system, uniform filter banks with 32 complex conjugated subbands and a decimation factor of 16 divide fullband signals $e(n)$ and $u(n)$ into subbands for the subband estimation. The subband NLMS step size is chosen to be $2^{-12}$ for all subbands except the lowest one (DC band), which is set to zero, so that the AFC is not performed under approximately 500 Hz. The subband NLMS regularization parameter is $2^{-16}$.

For the FS-AFC system, we use the same S-AFC system to estimate $\hat{h}(n)$. In addition, a frequency shifting of 10 Hz is performed in the forward path $f(n)$ for frequencies above 1000 Hz, because performing frequency shifting at the lowest frequencies would not be beneficial since the AFC system is already turned off, and it could significantly degrade sound quality, as shown in Section 2.

For the PN-AFC system, we again use the S-AFC system to estimate $\hat{h}(n)$. Furthermore, the probe noise signal is generated using a masking model, and the probe noise enhancement filter $a(n)$ is estimated using the same subband adaptive filter structure for estimating $\hat{h}(n)$, with the subband step size of $2^{-2}$ for all subbands.

3.3 Test Results and Discussions

In this section we evaluate the test results in terms of the maximum coefficient misalignment $\epsilon_F(n)$ and the maximum stable gain $M(n)$. We consider $\epsilon_F(n)$ in four equally divided frequency regions from 0 to 10 kHz.

3.3.1 Speech Signals

Fig. 7 shows test results for an example speech signal in terms of $\epsilon_F(n)$ and $M(n)$. As expected, $\epsilon_F(n)$ decreases and $M(n)$ increases over time as the AFC systems adaptively improve the feedback path estimate; it continues until the feedback path undergoes a momentary change after 30 s. The result of this is a mismatch between the true and

Table 5. Test signals used in simulation experiments.

<table>
<thead>
<tr>
<th>Speech</th>
<th>Music</th>
</tr>
</thead>
<tbody>
<tr>
<td>Danish Female Speech</td>
<td>Classic</td>
</tr>
<tr>
<td>Japanese Female Speech</td>
<td>Female Singing</td>
</tr>
<tr>
<td>Norwegian Male Speech</td>
<td>Violin + Female Singing</td>
</tr>
<tr>
<td>English Female Speech 1</td>
<td>Flute</td>
</tr>
<tr>
<td>English Female Speech 2</td>
<td>Organ</td>
</tr>
<tr>
<td>English Male Speech</td>
<td>Piano + Female Singing</td>
</tr>
</tbody>
</table>

Fig. 6. Measured acoustic feedback paths with and without a telephone placed next to the ear of a hearing aid user. (a) Impulse Response. (b) Magnitude Response.
estimated feedback paths, which is seen as the jump/drop in $\varepsilon_f(n)$ and $M(n)$, respectively.

We observed that for this example speech signal, the AFC cancellation performance in terms of $\varepsilon_f(n)$ and $M(n)$ are very similar for almost all AFC systems except the F-AFC, which performs poorly at the lower frequencies $0 - 2.5$ kHz, where the correlation in speech signals is generally strong. The consequence of this is that a relatively small $M(n)$ less than 35 dB (as an additional gain less than 5 dB added to the initial value of $M(0) \approx 30$ dB) will be available in the system, as seen in Fig. 7(b), even though the desired amplification in hearing aids can be more than 50 dB [31]. In this example, the improvement of the maximum stable gains can be more than 10 dB in the other AFC systems. Furthermore, Fig. 8 shows the averaged maximum stable gain obtained from all simulations of all speech signals. Again, it is seen that the F-AFC performs much worse than the other AFC systems, which have similar behaviors.

The PEM-AFC system works generally well in this case, since the assumption of autoregressive incoming signals $x(n)$ is reasonably valid especially for the unvoiced regions of speech signals [41]. The S-AFC system transforms the fullband signals $u(n)$ and $e(n)$ to subbands, this transform generally decorrelates the signals [1], and an improvement is thereby obtained. In the FS-AFC and PN-AFC systems, the frequency shifting and probe noise are applied in addition to the subband transform, it seems that any further improvement is very limited compared to the S-AFC system. Furthermore, the PEM-AFC and PN-AFC seem to have a slightly faster convergence but lower maximum stable gain in the steady-state with the chosen simulation parameters.

### 3.3.2 Music Signals

Fig. 9 shows similar simulation results for an example music signal. It is very clear that the F-AFC does not work properly, due to the much stronger correlation, over the entire frequency spectrum, in the music signal. Interestingly, performance of PEM-AFC and S-AFC is not much better than F-AFC in terms of MSG $M(n)$ and MCM $\varepsilon_f(n)$ below 5 kHz. The PEM-AFC system is not performing well due to the assumption of the incoming signals $x(n)$ to be autoregressive is violated in the music signal case, and the prefilters in the PEM-AFC system can thereby not decorrelate the loudspeaker signal $u(n)$ from the incoming signal $x(n)$. The subband transform in the S-AFC system still has a decorrelation effect, but it is no longer adequate.

Furthermore, it is clear that the FS-AFC and PN-AFC systems perform significantly better than other systems in this case. Fig. 9(a) shows that the FS-AFC system is very effective to provide unbiased estimation at higher frequencies above 5 kHz, but it has only limited effects in lower frequencies, due to the stronger correlation in signals. On the other hand, the PN-AFC is able to provide unbiased estimation in the entire frequency range, and it provides the highest maximum stable gain as seen in Fig. 9(b).

The same trend is observed in Fig. 10, which shows the averaged maximum stable gain obtained from all simulations of music signals. It is clear that the F-AFC system fails to increase the maximum stable gain over time, the PEM-AFC and S-AFC can only provide minimal increments, but the FS-AFC improves it by more than 6 dB, and the PN-AFC improves it by more than 12 dB.
Table 6. Slopes [dB/s] of average maximum stable gain curves.

<table>
<thead>
<tr>
<th></th>
<th>Speech</th>
<th>Music</th>
</tr>
</thead>
<tbody>
<tr>
<td>F-AFC</td>
<td>0.41</td>
<td>−0.25</td>
</tr>
<tr>
<td>PEM-AFC</td>
<td>0.75</td>
<td>0.11</td>
</tr>
<tr>
<td>S-AFC</td>
<td>0.73</td>
<td>0.22</td>
</tr>
<tr>
<td>FS-AFC</td>
<td>0.72</td>
<td>0.54</td>
</tr>
<tr>
<td>PN-AFC</td>
<td>0.83</td>
<td>1.23</td>
</tr>
</tbody>
</table>

3.3.3 Speech Versus Music Signals

To complete the observations, we computed the slopes of the average maximum stable gain curves based on the first five seconds of all simulation results. Table 6 shows the computed slopes. This data demonstrates that all AFC systems except the F-AFC perform well for speech signals, but only the FS-AFC and especially the PN-AFC are robust against music signals.

Furthermore, an interesting and somewhat unexpected observation is that the PEM-AFC has similar performance as the S-AFC system for speech signals, whereas the S-AFC system is slightly more robust against correlation in music signals (compared to this particular version of the PEM-AFC). A similar conclusion is found in a comparison between an DCT domain AFC system and a PEM based AFC system in a recent work [42].

4 COMPUTATIONAL COMPLEXITY EVALUATION

In addition to the sound quality and AFC performance improvement evaluations, another important consideration is the computational complexity and memory usage for each AFC system. In this section we make a rough complexity estimation of each system by counting the number of required real multiplications; we do not take into account that specific optimizations and modifications of algorithms could reduce the complexity significantly for a particular system, such as partly sharing the subband processing of the hearing aid forward path with subband AFC algorithms. Hence, our estimates should be considered as upper-bounds for the complexity.

We estimated the complexity for each AFC system in a two-microphone-channel hearing aid system. The exact number of multiplications are computed based on the chosen parameters for simulations in this work. For convenience, we normalized the results relative to the F-AFC reference system. Furthermore, we divided each AFC system into functional subgroups and computed the complexity in percentage for each subgroup. The results are given in Table 7.

The F-AFC system has the lowest computational complexity, the adaptive filter estimation is done cheaply using the NLMS algorithm. The PEM-AFC needs an additional prefilter estimation and the subsequent filtering on error and reference signals for the adaptive filter estimation. The S-AFC system is more complex due to the need of filter banks to divide fullband signals into subbands, where a subband NLMS algorithm works on complex signals rather than real signals in the F-AFC system, and the frequency domain feedback path estimate has to be transformed back to a time-domain filter. The FS-AFC needs an additional frequency shifting algorithm compared to the S-AFC system. Finally, the PN-AFC system needs probe noise generation, probe noise enhancement including filtering of error
that is robust in even the most challenging feedback situations. With the increasing computational power available in hearing aids, these improved cancellation systems can be realistically implemented in the near future.

6 REFERENCES


Table 7. Complexity factors (CF) for different AFC systems.

<table>
<thead>
<tr>
<th>System</th>
<th>CF</th>
<th>Subgroup Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>F-AFC</td>
<td>1</td>
<td>NLMS update 67%</td>
</tr>
<tr>
<td>PEM-AFC</td>
<td>1.27</td>
<td>Prefilter estimation 9%</td>
</tr>
<tr>
<td>S-AFC</td>
<td>1.55</td>
<td>Prefilter filtering 12%</td>
</tr>
<tr>
<td>FS-AFC</td>
<td>1.68</td>
<td>F-AFC System 79%</td>
</tr>
<tr>
<td>PN-AFC</td>
<td>2.78</td>
<td>Subband transform 12%</td>
</tr>
</tbody>
</table>

and reference signals similar to the PEM-AFC system, in addition to the subband adaptive filter estimation as in the S-AFC system.

Table 7 shows, perhaps as expected, that the systems with best AFC performance as demonstrated in Section 3 also have higher computational complexity.

5 CONCLUSION

We conducted an evaluation of several state-of-the-art acoustic feedback cancellation systems for hearing aids in terms of the cancellation performance, sound quality degradation, and computational complexity. In particular, we compared a traditional fullband system to a prediction error method based fullband system, a subband system, a subband system with frequency shifting, and a recently proposed subband system with a novel probe noise usage. By allowing a perceptually noticeable but small sound quality degradation in the loudspeaker signal to decorrelate it from the incoming signal, we evaluated the cancellation performance in terms of maximum coefficient misalignment of the adaptive filters and the maximum stable gain in hearing aid simulations. All systems outperformed the traditional fullband system in cancellation performance. Especially the subband system with probe noise provided an improvement of more than 12 dB in maximum stable gain in the most difficult situations, but it is also the most computationally complex system, roughly 2.8 times more complex than the traditional system. The subband system with frequency shifting improved the maximum stable gain by more than 6 dB with a complexity increment by a factor of 1.7 for the same situations. Furthermore, we showed that the subband system had a slightly larger improvement than the prediction error method based fullband system, which had the lowest computational complexity increment by a factor of 1.3 compared to the traditional system. In this way, the systems providing largest improvements are also more computationally complex. Hence, choosing an appropriate system for a practical application, among these evaluated, is a clear compromise between performance and computational cost. However, for a price of 2 to 3 times the reference complexity, a cancellation system can be realized


Meng Guo received the M.Sc. degree in applied mathematics from Technical University of Denmark, Lyngby, Denmark, in 2006, and the Ph.D. degree in signal processing from Aalborg University, Aalborg, Denmark, in 2013. From 2007 to 2010 he was with Oticon A/S, Smørum, Denmark, as a research and development engineer in the area of acoustic signal processing for hearing aid applications, especially in algorithm design of acoustic feedback cancellation. Currently, he is with Oticon A/S, as a research engineer. His main research interests are in the area of acoustic signal processing, including acoustic feedback cancellation, acoustic echo cancellation, adaptive filtering techniques, microphone array processing, and speech enhancement.

Søren Holdt Jensen received the M.Sc. degree in electrical engineering from Aalborg University, Aalborg, Denmark, in 1988, and the Ph.D. degree in signal processing from the Technical University of Denmark, Lyngby, Denmark, in 1995. Before joining the Department of Electronic Systems of Aalborg University, he was with the Telecommunications Laboratory of Telecom Denmark, Ltd., Copenhagen, Denmark; the Electronics Institute of the Technical University of Denmark; the Scientific Computing Group of Danish Computing Center for Research and Education (UNiX), Lyngby, Denmark; the Electrical Engineering Department of Katholieke Universiteit Leuven, Leuven, Belgium; and the Center for PersonKommunikation (CPK) of Aalborg University. He is full professor and is currently heading a research team working in the area of numerical algorithms, optimization, and statistical signal processing for speech and audio processing, image and video processing, multimedia technologies, and digital communications. Prof. Jensen was an associate editor for the IEEE Transactions on Signal Processing and Elsevier Signal Processing, and is currently associate editor for the IEEE Transactions on Audio, Speech and Language Processing and EURASIP Journal on Advances in Signal Processing. He is a recipient of an European Community Marie Curie Fellowship, former chairman of the IEEE Denmark Section, and founder and chairman of the IEEE Denmark Section’s Signal Processing Chapter. He is member of the Danish Academy of Technical Sciences and was in January 2011 appointed as member of the Danish Council for Independent Research—Technology and Production Sciences by the Danish Minister for Science, Technology, and Innovation.

Jesper Jensen received the M.Sc. degree in electrical engineering and the Ph.D. degree in signal processing from Aalborg University, Aalborg, Denmark, in 1996 and 2000, respectively. From 1996 to 2000, he was with the Center for PersonKommunikation (CPK), Aalborg University, as a Ph.D. student and Assistant Research Professor. From 2000 to 2007, he was a Post-Doctoral Researcher and Assistant Professor with Delft University of Technology, Delft, The Netherlands, and an External Associate Professor with Aalborg University. Currently, he is a Senior Researcher with Oticon A/S, Smørum, Denmark, where his main responsibility is scouting and development of new signal processing concepts for hearing aid applications. He is also a Professor with the Section for Multimedia Information and Signal Processing (MISP), Department of Electronic Systems at Aalborg University. His main interests are in the area of acoustic signal processing, including signal retrieval from noisy observations, coding, speech and audio modification and synthesis, intelligibility enhancement of speech signals, signal processing for hearing aid applications, and perceptual aspects of signal processing.