Study on the applicability of instrumental measures for black-box evaluation of static feedback control in hearing aids

N. Madhu[a] and J. Wouters
ExpORL–Department of Neurosciences, K.U. Leuven, B-3000 Leuven, Belgium

A. Spriet
ExpORL–Department Neurosciences and ESAT/SISTA–K.U. Leuven, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium

T. Bisitz
HörTech gGmbH, Marie-Curie-Straße 2, D-26129 Oldenburg, Germany

V. Hohmann
Department of Medical Physics, Carl von Ossietzky Universität Oldenburg, D-26111 Oldenburg, Germany

M. Moonen
ESAT/SISTA-K.U. Leuven, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium

(Received 13 July 2010; revised 23 March 2011; accepted 20 May 2011)

Presented is a report on black-box evaluation of feedback control systems for commercial hearing aids. The aim of the study is to examine the ability of existing instrumental measures to quantify the performance of the feedback control system in black-box settings and on realistic signals, when more than one element of the signal processing chain may be active (compression, noise suppression, microphone directionality, etc.). The evaluation is carried out on 6 different hearing aids and for 10 measures. Thereby it is possible to see which measure is best suited to measuring which specific characteristic of the feedback control system, and serves as a beginning for conducting perceptual tests. The study uses static (but variable) feedback paths and is based on signals recorded from the in-ear microphone of an artificial head, on which the hearing instruments are mounted.
© *2011 Acoustical Society of America*. [DOI: 10.1121/1.3605526]

PACS number(s): 43.66.Ts, 43.60.Qv [MAH] Pages: 933–947

I. INTRODUCTION

There is an increasing trend toward the use of hearing aids with open and vented couplings in hearing impaired subjects. Especially with such couplings, the closed-loop system from the receiver of the hearing aid to the microphone(s) could become unstable when a large forward path gain is introduced, leading to annoying acoustic feedback. This places a limitation on the maximum amplification the hearing aid can provide and, for users with moderate to severe hearing loss, this might not be enough. Feedback control systems[1–9] help reduce the effects of instability and are incorporated into hearing aids as a standard component in the signal processing chain. Further, the acoustic path between the receiver and the microphone(s) can vary significantly with time, depending on the acoustic environment [e.g., if one brings a telephone handset or the palm of a hand close to the ear, the feedback path attenuation can be temporarily reduced by 10 to 20 dB (Refs. 10–13)]. Therefore adaptive feedback control techniques need to be implemented to deal with such feedback path changes.

As the feedback control system is a critical element of the signal processing chain, development of robust procedures and measures to quantify its performance become imperative. Since a feedback control system basically permits an increase of the forward path gain without the system becoming unstable, the amount by which this gain can be increased characterizes one aspect of its performance. This increased forward path gain is typically measured in terms of the *maximum stable gain* (MSG),[2,6,14–16] or the loop gain.[17,18] Alternatively, measures for the detection of *instability* (and thereby, indirectly, the maximum gain) have been proposed by the authors of Refs. 14 and 19 and are based on the power concentration ratio (PCR) and the hearing aid transfer function variation (TVC), respectively. Computation of these measures is done on the basis of a spectrally flat input signal. However, signals encountered in practice are spectrally colored (e.g., music or speech). For such signals feedback control based on adaptive algorithms could encounter problems:[3,7,20] due to correlation between the input to the feedback cancellation algorithm and the hearing aid input signal, the adaptive system may erroneously attempt to cancel the *input* signal instead of the acoustic feedback. Consequently, the application of the above measures to a white noise input often yields an overestimation of the maximum gain that can be applied in real-life scenarios. Also, the MSG is typically derived based on the audiogram settings.[14,18] However, some devices apply *gain limitation* in order to avoid feedback. In this case, the gain setting will not correspond to the actually applied gain, which further complicates the analysis.

[a]Author to whom correspondence should be addressed. Electronic mail: nilesh.madhu@rub.de

Downloaded 12 Mar 2012 to 134.58.253.57. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp

In addition to providing a high MSG, the feedback control technique should preserve a good sound quality. Signal fidelity is thus another performance aspect. Means to estimate the distortion introduced by feedback reduction algorithms have been proposed in Refs. 14 and 18. Whereas Ref. 14 investigated the susceptibility to pure tones, Ref. 18 assessed the robustness of the feedback reduction system to periodic input signals.

Generally the evaluation procedures assume that the hearing aid functions as a linear amplifier. However this is not the case in practice due to non-linear processing such as gain limitation, saturation of the receiver at high gains, dynamic range compression, noise suppression, etc. While these effects may be neglected if one has access to the feedback canceller input and output, in black-box settings, one can only observe the hearing aid input and output. Therefore, the evaluation procedures need to be modified taking this into account.

In addition, most measures require a feedback free reference signal against which the hearing aid output is computed. This is typically obtained as the hearing aid output recorded at a gain far below instability. However, the underlying assumption here is that the hearing aid functions linearly. Therefore an alternative procedure is used for estimating the feedback-free reference signal for this case, based on replacing an open fitting by a closed fitting, with appropriate compensation.[21]

These limitations of existing instrumental measures are addressed in this contribution. Specifically, we perform black-box evaluation of the feedback cancellers of commercial hearing aids where different elements in the signal processing chain could be active. The performance evaluation is done for six different commercial hearing aids of different manufacturers, on the basis of the existing instrumental measures—suitably modified for the black-box case and on a wide variety of input signals.

We stress that it is not the aim of this study to present a *comparison* of the feedback control systems of different manufacturers. Rather, we have intentionally selected a wide range of settings and hearing aids in order to assess the *general applicability* of existing measures for feedback control system evaluation. Some of the measures were already tested in Ref. 21 but for the case where the hearing aid settings and gains were matched, the aim being to specifically *compare* algorithms, whereas here we see which measures are applicable to testing the different aspects of feedback control system performance in black-box settings.

The next stage would be to identify a subset of measures that perform robustly in such settings. Once such instrumental measures are obtained, the application of quality models[22–26] and the correlation with perceived quality can be studied.

II. EXPERIMENTAL SETUP, HEARING AIDS AND STIMULI

A. Experimental setup

The hearing aids were fitted with an open coupling realized using a Phonak Fit-and-Go open fit (cf. Ref. 27), and mounted on the left ear of a Cortex II artificial head. The performance measures are computed *offline* based on the signals recorded at the left in-ear microphone (at 0 dB amplification) in response to the presented stimulus. The recordings were performed in a room of dimensions (5.6 m × 3 m × 2.4 m), and an average reverberation time (T_{60}) of 0.13 s (for frequencies between 250 Hz and 6.3 kHz). The background noise level was around 30 dBA. The artificial head was placed on a fixed-mount to prevent any unwanted change in position during insertion and removal of the hearing aids and couplings.

The stimuli were presented through a Genelec loudspeaker positioned at a distance of 1 m in front of the center of the head. Playback and recording was done in a synchronous manner using an RME-Multiface A/D-converter, connected to an RME Hammerfall DSP card on a standard PC running Windows XP. For controlling playback and recording of the sound signals, MATLAB with SOUNDMEXPRO was used, which allows synchronous playback and recording with a constant delay/latency via the ASIO driver of the soundcard. The automation of the recordings was done using the freeware scripting tool AUTOIT[28] along with the fitting tool provided by each manufacturer. Whereas the recordings have been done at a sampling frequency of 44.1 kHz, the signals are downsampled to 16 kHz before computing the measures.

While the setup allows dynamic feedback paths, this was not used in this study of performance measure comparison. Two static acoustic conditions were tested, referred to as "no obstruction" and "board." In the latter condition, an aluminum board of dimensions (20 cm × 20 cm × 3 mm) was positioned at a distance of 2 cm from the pinna of the left ear. To ensure the reproducibility of this setting, a servo-controlled linear-motor assembly was used with the head and the linear motor being connected to a common metal plate. This is illustrated in Fig. 1. In the "no obstruction" condition, there was no foreign object in the vicinity of the head and the motor setup was completely removed.

The position of the artificial head and loudspeaker were fixed for the complete duration of the recordings. Therefore significant problems with reproducibility are not expected.

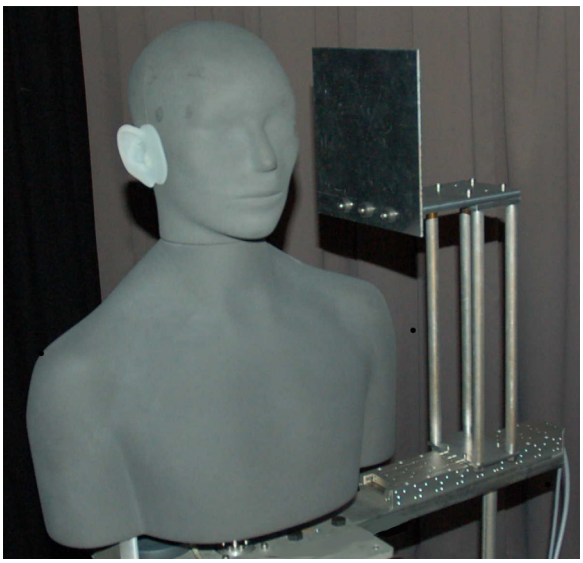

FIG. 1. (Color online) Setup with board.

Downloaded 12 Mar 2012 to 134.58.253.57. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp

TABLE I. Overview of the hearing aids used.

Aid-identifier	Type	Open coupling	Settings
A	BTE	Phonak fit and go	Mic. omni, noise suppression off, gain limitation off. Feedback control through notch filtering.
B	RITC/BTE	open dome (on receiver)	Mic. omni, noise suppression off, gain limitation off. Feedback control through forward path estimation and cancellation. Phase modifications are introduced into the signal to improve convergence for correlated input.
E	BTE	Phonak fit and go	Mic. omni, noise suppression off, gain limitation on. Adaptive feedback cancellation system.
F	BTE	Phonak fit and go	Mic. directive, noise suppression on, gain limitation on. Adaptive feedback cancellation system.
H	BTE	Phonak fit and go	Mic. omni, noise suppression off, gain limitation on. Adaptive feedback cancellation system.
J	BTE	Phonak fit and go	Mic. directive, noise suppression on, gain limitation on. Adaptive feedback cancellation system.

This was checked by comparing the acoustic paths from the loudspeaker to the in-ear microphone on a regular basis—where very minor differences were found.

B. Hearing aids and settings

Hearing aids provided by Siemens, Phonak, GNResound, Oticon, and Widex were used in this study. For the evaluation, the settings for each hearing aid were fixed following the information given by the manufacturers regarding the optimal fitting. Table I gives an overview of their types and settings. In devices A and B, gain limitation and noise suppression were disabled. In devices E and H noise suppression and directionality were disabled. In devices F and J, all additional processing that would be generally used in daily life was enabled.

C. Stimuli

A wide range of input stimuli, with different spectral characteristics, were used in this study. Table II shows a list of all signals and their corresponding presentation levels. Each stimulus was 60 s in length, with a 5 s period of silence before and after presentation. For level calibration, the free-field level in the absence of the head and at its center position was measured using a broadband, speech-shaped noise signal. All other signals were calibrated with respect to this signal by comparing the A-weighted rms values of the signals.

III. SIGNAL MODEL

A. Notation

The in-ear microphone signal is denoted as $y[k]$, and consists of two components: the hearing aid output $u[k]$ and the sound that directly enters the ear through the vent of the earmold, i.e., the direct signal path component $d[k]$. The latter is measured by recording the in-ear microphone signal with the hearing aid switched off. The hearing aid output $u[k]$ can then be computed as

$$u[k] = y[k] - d[k], \tag{1}$$

as the acquisition of both $d[k]$ and $y[k]$ is synchronized with the presented stimulus.

To take into account spectral coloration of the input signal the actual hearing aid output $u[k]$ is compared with the hearing aid output that would have been obtained in the *absence* of feedback, for the same acoustic scenario and hearing aid settings (cf. Sec. III B). This signal forms the reference, and is denoted as $r_u[k]$. Similarly, the reference for the in-ear signal $y[k]$ is denoted by $r_y[k]$.

Some measures are computed on the basis of the short-time power spectral density (PSD) of the signals. The PSD of a signal $u[k]$ is denoted as $P_u(f, l)$, where l is the time-frame index and f is the frequency index.

TABLE II. List of all test signals and playback levels.

Signal	Level(s) (dB SPL)	Comment
White noise	65, 85	Bandlimited to 10 kHz
Stationary, speech-shaped noise	65	From the HINT (Ref. 40) database
Sentences from a male speaker	65	From the HINT database
ISTS	55,65,80	International speech test signal (Refs. 41–43)
Impulsive sounds	peak value of 100	Opening and closing window, keys falling
Organ	65	Recording of organ music
Chamber music	65	Recording of small chamber music ensemble
Opera	65	"Der Hölle Rache" from "Die Zauberflöte" by Mozart

Downloaded 12 Mar 2012 to 134.58.253.57. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp

B. Estimation of the feedback-free reference

In black-box settings, where there is no access to the hearing aid microphone and receiver signal, the hearing aid output (as measured by the in-ear microphone) at a gain far below instability and with the feedback canceller disabled is typically used as a reference signal.[18,19] The gain difference between the actual hearing aid output and the low-gain output is then compensated for. This procedure, however, assumes that the hearing aid behaves as a linear system—an assumption that does not hold in general.

Therefore an alternative procedure, proposed in,[21] is used for estimating the reference signal, that is, applicable also when the system is non-linear. The in-ear signal $y_{\text{closed}}[k]$ with the feedback canceller *disabled* is recorded at the same hearing aid settings as the actual in-ear signal $y[k]$, but with a *closed* instead of an open fitting.[21,29] In addition, the direct path component $d_{\text{closed}}[k]$ with the closed-fitting was measured as the in-ear signal with the hearing aid switched off. For the closed fitting, a fully occluded, hard earmold was used. The difference in frequency characteristic of the ear canal due to the closed fitting is compensated for by means of a finite impulse response filter $w[k]$. For the receiver-in-the-ear, the ear canal was sealed by putting foam [from a temporary foam earmold E-A-RTEMP 13A (EARtone)] around the receiver.

The reference signals $r_u[k]$ and $r_y[k]$ were then computed as follows.

(1) Compute the hearing aid output $u[k]$ with the open fit and the hearing aid output $u_{\text{closed}}[k]$ with the closed fit:

$$u_{\text{closed}}[k] = y_{\text{closed}}[k] - d_{\text{closed}}[k]. \tag{2}$$

(2) Estimate the compensation filter $w[k]$ as the L_w-tap filter that minimizes the mean square error between $u_w[k]$ and $w[k] * u_{w,\text{closed}}[k]$, where $u_w[k]$ and $u_{w,\text{closed}}[k]$ are the hearing aid outputs at a gain far below instability and $*$ is the convolution operator. For the identification of $w[k]$, a white noise input signal with a presentation level of 85 dB sound pressure level (SPL) was used, with the hearing loss selected such that the hearing aid output is sufficiently above the environmental and internal noise level and yet sufficiently below instability.

(3) The reference signals $r_u[k]$ and $r_y[k]$ are then obtained as

$$\begin{aligned} r_u[k] &= w[k] * u_{\text{closed}}[k] \\ r_y[k] &= w[k] * u_{\text{closed}}[k] + d[k]. \end{aligned} \tag{3}$$

C. Additional considerations

The environmental noise $n[k]$ was measured in the 5 s interval before and after stimulus presentation. A rough estimate of the environmental and internal noise component in $r_y[k]$ was obtained as the sum of the compensated closed-fit recording of silence and the open-fit recording of silence with the hearing aid switched off.

IV. PERFORMANCE MEASURES

This section briefly summarizes the measures that have been applied. The measures are classified into (1) broadband energy measures, (2) measures for detecting oscillations, (3) signal distortion measures, (4) perceptually motivated measures, and (5) gain measures.

For non-stationary signals (such as the speech and the opera signal), some segments may be more prone to feedback and oscillations than other segments. Therefore, unless otherwise mentioned, the measures will be computed using frames of 0.5 s with an overlap of 80% between adjacent frames. The PSD is then computed as follows: each frame of 0.5 s length is segmented into sections of 256 samples at 50% overlap and windowed with a von Hann window before being transformed to the DFT domain. The short-term PSD is then obtained for that frame as the averaged square amplitude of the DFT coefficients.

Further, to reduce the effect of differences in the environmental and internal noise component in $u[k]$ and $r_u[k]$, the measures are only computed for frames where the short-term energy of the reference signal $r_u[k]$ exceeds the environmental noise $n[k]$ by at least 10 dB.

A. Broadband energy measures

Instability results in a large increase in output signal level compared to the level measured in the absence of feedback. Broadband energy measures provide a good means to capture this behavior. Possible measures to detect instability and assess the amount of feedback are the output-to-reference signal energy ratio (ORR) and the feedback-to-reference signal (FSR) energy ratio.

1. Output-to-reference signal energy ratio

A simple method to detect instability is to track the short-term output-to-reference signal energy ratio ORR(l)[15]

$$\text{ORR}(l) = 10\log_{10}\frac{\sum_{i=k-L/2}^{k+L/2} u^2[i]}{\sum_{i=k-L/2}^{k+L/2} r_u^2[i]}. \tag{4}$$

$L+1$ is the window length used in the energy computation where L was set to 8000 samples at 16 kHz. Instability is said to occur if the energy ratio exceeds a certain threshold, e.g., 6 dB. The energy ratio ORR(l) detects an increase in the output signal level caused by feedback. However, below instability, it does not give any information about the amount of residual feedback.

2. Feedback-to-reference signal energy ratio

To quantify the amount of feedback, the short-term feedback-to-reference signal energy ratio FSR(l) is computed as[16,29,30]

$$\text{FSR}(l) = 10\log_{10}\frac{\sum_{i=k-l/2}^{k+L/2} (u[i] - r_u[i])^2}{\sum_{i=k-L/2}^{k+L/2} r_u^2[i]}. \tag{5}$$

In addition, the short-term intelligibility weighted feedback-to-reference signal energy ratio $\text{FSR}_{\text{intellig}}(l)$ in the hearing aid output is computed as[29]

Downloaded 12 Mar 2012 to 134.58.253.57. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp

$$\mathrm{FSR}_{\mathrm{intellig}}(l) = \sum_i w_{\mathrm{ERB}}[i]\mathrm{FSR}_i, \tag{6}$$

where FSR_i is the feedback-to-reference signal energy ratio in the ith auditory critical band, defined by the equivalent rectangular bandwidth (ERB) of auditory filters[31]

$$\mathrm{FSR}_i = 10\log_{10}\frac{\max\left(\int_{f\in B_i} P_v(f)df, \beta\int_{f\in B_i} P_n(f)df\right)}{\max\left(\int_{f\in B_i} P_{r_u}(f)df, \beta\int_{f\in B_i} P_n(f)df\right)}, \tag{7}$$

with $P_v(f, l)$ being short-term PSD of the feedback component $v[k] = u[k] - r_u[k]$, and $P_n(f)$ the *long-term* PSD of the environmental noise $n[k]$. The weight $w_{\mathrm{ERB}}[i]$ gives an equal weight to each auditory critical band B_i between 300 and 6500 Hz.

To limit the impact of environmental and internal noise on the measurements, the feedback and the reference signal energy within each band are lower bounded by β times the noise energy (with $\beta = 2$). Since the bandwidth of the auditory critical bands increases with frequency, high frequency oscillations are weighted less in $\mathrm{FSR}_{\mathrm{intellig}}$ than low frequency oscillations.

B. Measures for detecting oscillations

Even before instability occurs, oscillations can already be perceived, and can be measured by the power concentration ratio (PCR) and the hearing aid transfer function variation criterion (TVC), respectively. Originally (cf. Refs. 14 and 19) these criteria were only defined for a white noise input signal. In addition, the PCR depends on the hearing aid frequency response. Below, these measures are modified so that they can be applied to spectrally colored input signals and are robust to spectral peaks in the hearing aid frequency response.[21,29,30]

1. Transfer function variation criterion

In Ref. 19 the hearing aid transfer function is measured for increasing gain values and a white noise input signal. The hearing aid transfer function in the initial stable condition acts as the reference hearing aid transfer function. This reference transfer function is multiplied by the gain difference between the increased gain and the stable gain. The difference in dB between the amplitude characteristic of the hearing aid transfer function and the gain compensated reference hearing aid transfer function is computed and is referred to as the transfer function variation function (TVF). The transfer function variation criterion (TVC) is then defined as

$$\mathrm{TVC}(l) = \max_f(|\mathrm{TVF}(f, l)|). \tag{8}$$

For spectrally colored input signals, an estimate of TVF(f, l) is obtained as

$$\mathrm{TVF}(f, l) = 10\log_{10}\left(\frac{P_u(f, l)}{P_{r_u}(f, l)}\right), \tag{9}$$

which may then be used in Eq. (8) (cf. Refs. 21, 29, and 30).

To reduce the influence of the noise component $P_u(f, l)$ and $P_{r_u}(f, l)$ are constrained such that only frequencies where the energy exceeds the environmental noise energy by a preset threshold are taken into account.

The TVC detects the largest peak or dip in the transfer function variation function. Instability is assumed when TVC > 20 dB.[29] However, it does not take into account the signal power $P_u(f, l)$ at the detected oscillation frequency. For non-white input signals, not all frequencies are equally excited. As a result, the detected oscillation frequency may be masked by non-critical frequencies with more power.

2. Power concentration ratio

The power concentration ratio (PCR) is defined in Ref. 14 as the degree to which a large amount of power is concentrated at a small number of frequencies in the hearing aid output. The measure, however, assumes that the hearing aid input signal is white. For spectrally colored input signals, two modified measures based on the PCR have been presented in Ref. 21. The oscillation frequencies are determined based on the transfer function variation function. Essentially these are the frequencies with a TVF(f, l) $\geq$ 6 dB. The selected frequencies are limited to a maximum of five, with the selection corresponding to those frequencies with the largest TVF(f, l).

The first measure is referred to as the *difference* in power concentration ratio, $\Delta\mathrm{PCR}(l)$, between the hearing aid output and the reference signal[21,30] and is defined as

$$\Delta\mathrm{PCR}(l) = \mathrm{PCR}_u(l) - \mathrm{PCR}_{r_u}(l). \tag{10}$$

Caution should be used when the power concentration $\mathrm{PCR}_{r_u}(l)$ of the reference signal approaches 1 (i.e., when the reference signal has most of its energy at the oscillation frequencies). In this case, $\Delta\mathrm{PCR}(l)$ approaches zero, even when oscillations occur. Further, note that oscillation frequencies with a large signal power $P_u(f, l)$ result in a larger difference $\Delta\mathrm{PCR}(l)$ than oscillation frequencies with a low signal power.

A second measure is referred to as the *normalized* power concentration ratio $\mathrm{PCR}_{\mathrm{norm}}(l)$ and consists in computing the traditional PCR on the normalized $P_u(f, l)/P_{r_u}(f, l)$. The fraction of the normalized power $P_u(f, l)/P_{r_u}(f, l)$ that is located at the five (or less) strongest oscillation frequencies is referred to as $\mathrm{PCR}_{\mathrm{norm}}(l)$. The normalized power concentration ratio does not take into account the power of the oscillation frequencies, i.e., all frequencies are equally weighted.

The modified TVC and PCR measures are computed on the frequency range from 650 to 6500 Hz. Outside this frequency range, the hearing aid output power for an open fitting is low and hence, susceptible to noise.

C. Spectral signal distortion measure: Frequency-weighted log-spectral signal distortion

To assess the distortion of the hearing aid output $u[k]$, the frequency-weighted log-spectral signal distortion SD(l) is defined as

$$\mathrm{SD}(l) = \sqrt{\int_{f_t}^{f_u} w_{\mathrm{ERB}}(f)\left(10\log_{10}\frac{P_u(f, l)}{P_{r_u}(f, l)}\right)^2 df}. \tag{11}$$

Downloaded 12 Mar 2012 to 134.58.253.57. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp

The frequency-weighting factor $w_{\mathrm{ERB}}(f)$ gives equal weight to each auditory critical band between $f_l = 300$ Hz and $f_u = 6500$ Hz. The short-term PSD $P_u(f, l)$ and $P_{r_u}(f, l)$ are constrained so that only frequency bins with energy level above the environmental and internal noise level are taken into account. This is a modified version of the standard log-spectral signal distortion measure [Ref. 32 (Chap. 3, p. 69)].

D. Perceptually motivated measure: Entrainment

When a feedback canceller is presented with a spectrally colored input signal, distortion artifacts may be generated (such as addition of tones, modulation-type distortion, etc.). This is referred to as entrainment.[18]

In Refs. 8 and 33 a perceptual correlate of the entrainment is calculated, based on the loudness and the sharpness of the entrained signal. The entrained signal is obtained by a modified spectral subtraction approach, where a short-term spectral gain is computed based on the difference between the short-term spectra of $y[k]$ and $r_y[k]$. This gain is applied to the spectrum of $y[k]$ which, upon resynthesis, yields a signal where the entrained segments are enhanced. The procedure for quantifying the entrainment is detailed below.

(1) Estimate the power spectral density (PSD) of $r_y[k]$, $y[k]$, and the background noise $n[k]$. Denote the respective PSD by $P_{r_y}(f, l)$, $P_y(f, l)$, and $P_n(f, l)$. The properties of auditory masking are taken into account when computing P_{r_y}.
(2) For computing the spectral subtraction gain, an "interference" signal is defined as

$$i[k] \triangleq r_y[k] + n[k], \tag{12}$$

i.e., the sum of the reference signal and the background noise. Assuming that $r_y[\kappa]$ and $n[k]$ are uncorrelated, the PSD of the interference signal defined as above may be written as

$$P_i(f, l) = p_{r_y}(f, l) + P_n(f, l), \tag{13}$$

where the individual components on the right hand side of Eq. (13) are separately computed as mentioned previously. The spectral subtraction gain $\mathcal{G}(f, l)$ is now computed based on this definition of the interference signal as

$$P_i(f, l) = \gamma P_{r_y}(f, l) + \gamma_n P_n(f, l), \tag{14}$$

$$\mathcal{G}(f, l) = \frac{P_y(f, l) - P_i(f, l)}{P_y(f, l)}, \tag{15}$$

where the γ and γ_n are oversubtraction factors. The values for these were selected as 1.25 and 3, respectively, in accordance with Ref. 33.
(3) Compute the *entrained* signal spectrum as

$$E(f, l) = \mathcal{G}(f, l) Y(f, l), \tag{16}$$

from which the corresponding signal $e[k]$ can be synthesized.
(4) Define the entrainment factor $\mathcal{H}(f, l) \triangleq (1 - \mathcal{G}(f, l))^{-1}$.
(5) Compute the specific loudness spLoud (z, l) and sharpness $\mathcal{S}(l)$ of $e[k]$ as described in Ref. 34. Note that spLoud is defined in the Bark scale[35] (and denoted by the variable z).
(6) Compute a weight factor $\mathcal{W}(z, l)$ in the Bark scale, based on the specific loudness and sharpness as

$$\mathcal{W}(z, l) = \frac{\mathrm{spLoud}(z, l)(1 + \max(\mathcal{S}(l) - 1.75, 0)}{25}. \tag{17}$$

(7) Finally, to obtain the entrainment level $\mathcal{L}_{\mathcal{H}}$, compute the product of $\mathcal{H}$ interpolated to the bark scale and the weighting factor $\mathcal{W}$. The maximum of the product for a frame is then taken as the $\mathcal{L}_{\mathcal{H}}$ for that frame, with values above 5 dB being considered as annoying.

The original approach in Ref. 33 is carried out with the hearing aid programmed in the linear mode (flat gain settings, no compression, noise suppression or microphone directivity) and below instability. Further, the reference signal is taken as the in-ear microphone signal with the feedback canceller off (obtained with the same fit as $y[\kappa]$).

However, for high gains, this signal cannot be used as the reference, as this setting might be unstable, leading to howling. Therefore, we propose to compute the entrainment in the following manner.

(1) Define the reference signal $r_y[k]$ as the signal obtained from the in-ear microphone with the *closed fit*, and transformed into the equivalent open-fit version as described in Sec. III B.
(2) Compute the entrainment level $\mathcal{L}_{\mathcal{H}}$ on the signals recorded on the open-fit and with the FBC on ($y_{\mathrm{on}}[k]$) and with FBC off ($y_{\mathrm{off}}[k]$), with respect to the reference signal $r_y[k]$.
(3) The *difference* between $\mathcal{L}_{\mathcal{H}\mathrm{on}}$ and $\mathcal{L}_{\mathcal{H}\mathrm{off}}$ then indicates the effect of the feedback canceller.

E. Gain measures

Some hearing aids sacrifice hearing aid gain and hence audibility to avoid feedback. As a result, also level or gain measures are needed.

1. Measurement of the power spectral density

The power spectral density (PSD) of the hearing aid output for a given test signal gives information about the output level and the behavior of the hearing aid as a function of hearing loss.

In addition, it allows to detect gain limitations by the hearing aid. By considering the PSDs for different test signal levels, information on compression can also be obtained.

The PSD is computed based on the hearing aid output signal $u[k]$ (both with feedback canceller on and off) as well as the reference signal $r_u[k]$. The PSD of the reference signal $r_u[k]$ represents the PSD that would be obtained in the absence of feedback.

The PSD of the output signal $u[k]$ can be compared with the reference PSD in order to detect oscillation frequencies

Downloaded 12 Mar 2012 to 134.58.253.57. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp

and other adaptive gain limitation methods (such as notch filtering).

2. Dip bandwidth

To measure reduction in gain by the feedback reduction technique in specific frequency regions, the dip bandwidth was proposed by Ref. 14. The dip bandwidth is defined as the bandwidth (in octaves) of the widest set of contiguous frequencies at which the gain deviation was below –5 dB. In, Ref. 14 a flat linear gain was assumed and the gain deviation was measured with respect to the hearing aid output with the feedback canceller disabled and at a gain below instability.

In this contribution, however, the gain deviation was computed with respect to the reference signal $r_u[k]$ instead of the hearing aid output with feedback canceller disabled. The gain deviation was obtained as

$$\text{Gain Deviation}(f,l) = \frac{P_u(f,l)}{P_{r_u}(f,l)}. \tag{18}$$

Note that this measure is affected by noise or estimation errors in reference signal and is restricted to frequencies with a high hearing aid output power.

V. MEASUREMENT PROCEDURE

The output of the hearing aid was recorded for different *fitted* hearing losses, from mild to severe. The template audiogram selected (corresponds to an ISMADHA standard loss) was an average audiogram for a moderate hearing loss (see Refs. 36 and 37), which was shifted in steps of 5 dB resulting in 20 different gain settings. The hearing-loss curve is illustrated in Fig. 2 (data in bold). The data in gray color shows the lowest and highest hearing loss curve that was used.

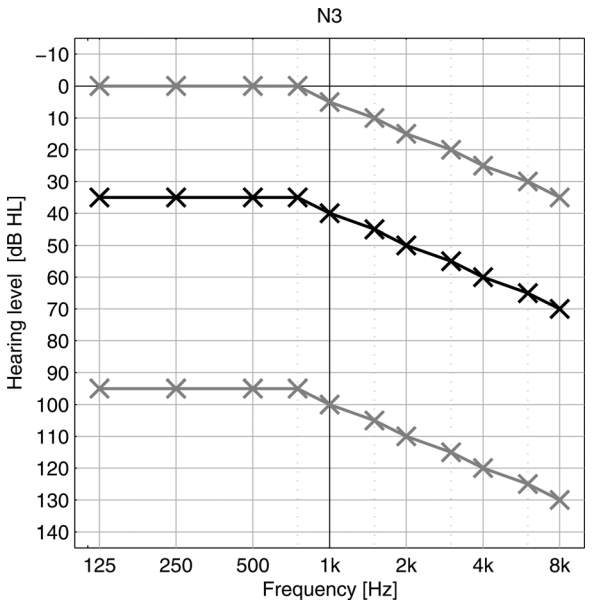

FIG. 2. Audiograms used for the evaluation, shifted in steps of 5 dB. The curve in black depicts the template audiogram.

For each audiogram setting, hearing aid and acoustic scenario (i.e., "no obstruction" or "board"), the following measurement procedure was adopted.

(1) Begin with the open fit and program the hearing aid using the corresponding audiogram settings and manufacturer recommendations for additional options.
(2) Present each stimulus and record the in-ear microphone signal with the hearing aid on. Additionally, for each stimulus, record the in-ear microphone signal with the hearing aid switched off and the in-ear microphone signal with the hearing aid switched on and the feedback canceller *disabled*.
(3) Replace the open fitting with the closed fitting and for each stimulus, record the in-ear microphone signal with the hearing aid switched off and the in-ear microphone signal with the hearing aid switched on and the feedback canceller disabled.

VI. RESULTS

We summarize here the advantages, disadvantages and applicability of each measure, and, unless specified otherwise, include the performance for one stimulus, the organ concert fragment; and one condition, that with no obstruction, to illustrate our points. We chose this stimulus because it is highly tonal and correlated, which is a difficult scenario for adaptive feedback control algorithms. We also selected the no obstruction case as the presence of the board did not yield increased amplification in the feedback path. In fact, it actually suppressed some frequencies. To illustrate the worst performance over the entire signal length, the 90 percentile of the frame-wise computed values of the measures ORR, FSR, $\text{FSR}_{\text{intellig}}$, TVC, ΔPCR, $\text{PCR}_{\text{fcnnorm}}$, and SD are considered. For the entrainment, the 99 percentile was taken, in accordance with the suggestion in Ref. 33.

Our observation in general was that stimuli such as impulsive sounds, ISTS speech at low level (55 dB SPL) and tonal signals such as the organ fragment constituted the most challenging conditions for the feedback control system. While we restrict ourselves to only one graphical example due to space constraints, the measure performance summary is based on the results obtained over all conditions and all stimuli.

For each device and measure, two curves are plotted. The dotted curve indicates the performance of the device with the feedback control system *disabled* whereas the dash-dotted curve indicates the performance with feedback control enabled. We reiterate that the graphical results presented are only illustratory, and cannot capture all the aspects of a particular measure.

A. Broadband energy measures

1. ORR

Figure 3 indicates the performance of the devices with respect to the ORR measure. With feedback control disabled, the presence of instability is clearly visible for devices A and B for high hearing losses. The other devices have gain limitation turned on, hence they prevent strong feedback, even

Downloaded 12 Mar 2012 to 134.58.253.57. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp

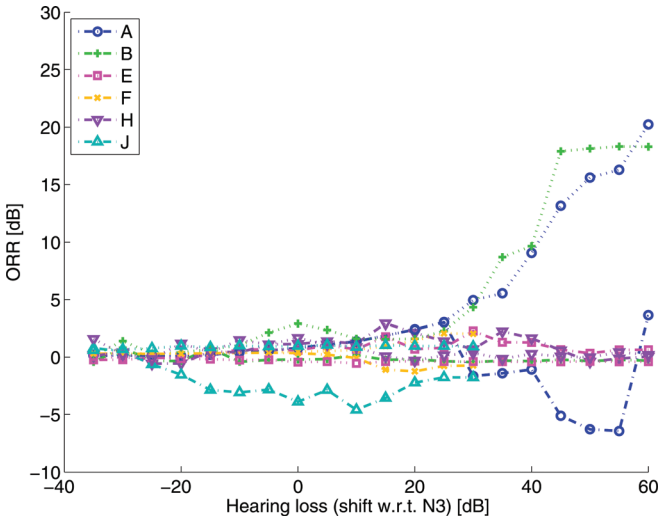

FIG. 3. Performance on the ORR measure for the organ stimulus. A value of over 6 dB is considered to correspond to instability.

when the feedback control system is disabled. With feedback control enabled, all devices do not show *strong* instability. Additionally, in this case, for hearing aids A and J, *negative* ORR values occur when feedback control is enabled. This is a sign that adaptive gain reduction is being applied (e.g., notch filtering for device A and gain limitation and noise suppression in device J).

ORR is applicable to all hearing aids and test signals. It is not only capable of detecting strong feedback (instability) but also indicates gain reduction. However, as it is a broadband measure, it is dominated by frequencies with the most energy. As a result, only strong feedback is well detected and weak oscillations are not always detected. This may be observed from the relatively flat values of ORR for lower hearing losses when the feedback control system activated and when it is disabled. This is in contrast to other measures such as PCR, TVC, and entrainment level, where the difference between the measures when the feedback control system enabled and when it is disabled is larger, even for low hearing losses.

2. FSR

Figures 4(a) and 4(b) depict the device behavior with respect to the FSR and the intelligibility-weighted FSR, respectively. In comparison to ORR, the measures gradually increase in value with increasing hearing loss. The effect of the feedback control system is clearly visible here: for the devices A and F, the curves for hearing aids with feedback control disabled have larger values than when feedback control is enabled; however, for devices E, H, and J, the measure values with feedback control disabled are *better* than when feedback control is enabled. For E and H this might indicate that the feedback control system actually degrades the output in this challenging condition with a tonal signal. For J, there is additional processing active, which could affect the output, in addition to the feedback control.

An exception is device B, which has constant, high values of FSR when feedback control is enabled. This constant

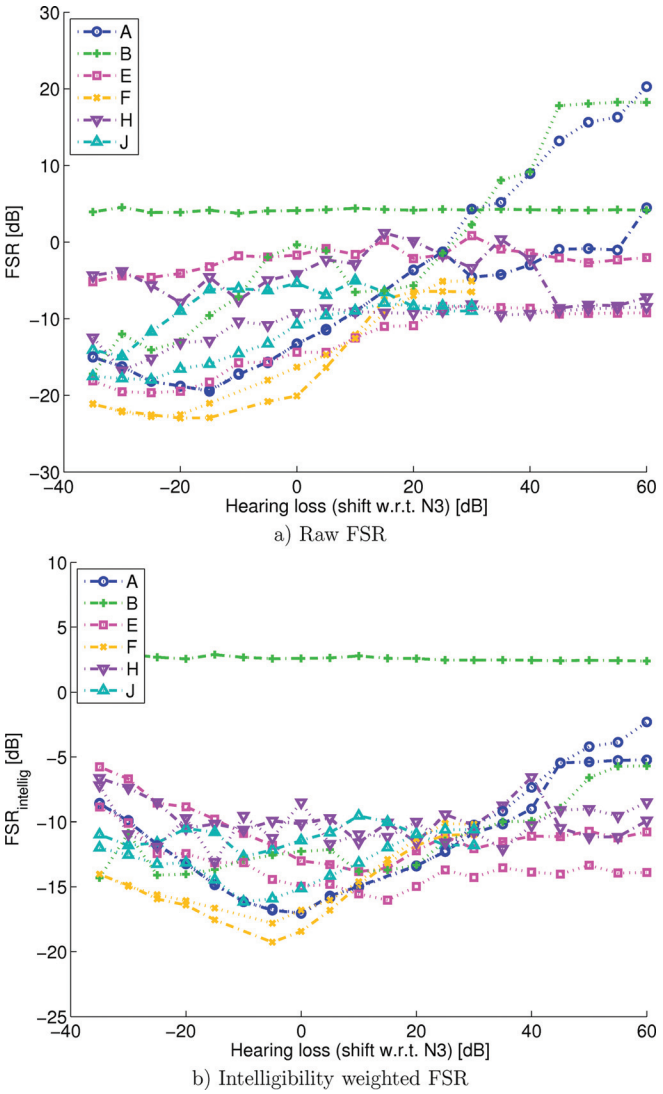

FIG. 4. Performance on the FSR measure for the organ stimulus. Values over 0 dB indicate instability.

high value is because the feedback control system introduces phase modifications in the forward path. The FSR relies on a successful subtraction of the reference signal for a reliable estimate. When the forward path contains gain and/or phase modifications, this is not reflected in the reference signal, yielding high values in the numerator in Eq. (5), with a corresponding high value of FSR.

FSR provides information on the amount of feedback (also below instability) and gradually increases with increasing feedback. It is sensitive to modifications introduced by the feedback control system. Therefore, it can detect phase modifications in the forward path due to the feedback canceller. However, in such cases where the feedback control system introduces any modification of the signal processing path, the measure is biased and does not give a reliable estimate of the residual feedback. In general, while such forward path modification yields biased values for most instrumental measures, FSR is sensitive to this.

The intelligibility-weighted FSR measure FSR$_{intellig}$ performs in a manner similar to FSR. However, as it applies frequency weighting according to intelligibility, higher,

Downloaded 12 Mar 2012 to 134.58.253.57. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp

narrowband frequencies are assigned a lower weight. However, as feedback most often occurs at higher frequencies, this measure is less sensitive as compared to FSR in this case. For the same reason, at low hearing losses, this measure is also more sensitive to noise (evidenced by the initial higher values, which first *reduce* with increasing hearing loss, before increasing again). This sensitivity comes about as, at lower hearing losses, noise dominates in the lower frequencies, which receive a higher weight.

B. Measures for detecting oscillations

1. TVC

The TVC criterion behavior is given in Fig. 5. Large values occur when the system is unstable (devices A, B for feedback control disabled). However, also below instability oscillations are detected, manifested by large values of TVC (e.g., devices E and H). The large values for device A at higher hearing losses, with feedback control enabled, is a result of the notch filtering.

Due to the normalization by the reference signal, the TVC measures the largest peak or dip in the transfer function difference independently of the signal power at that peak or dip frequency. Thus, even variations at low energy could result in a large TVC value. Thereby this measure can indicate both instability and low-energy oscillations. However, as dips may also be caused by gain reduction algorithms (e.g., notch filtering of device A), a distinction between feedback and gain reduction is difficult. The TVC may be modified to detect feedback only by considering just the peaks. In such a case, for device A with feedback control enabled, the large values for TVC will not occur.

The TVC is sensitive to errors in the hearing aid transfer function variation that are due to a low signal-to-noise ratio (as is the case at very low hearing losses) or an erroneous estimation of the reference signal, e.g., devices E and H have low output power above 5.5 kHz, and the compensation model for the creation of the reference signals for this device is inaccurate at frequencies above 5.5 kHz, resulting in some of the high TVC values for these devices. To prevent such biased results when using TVC, only the frequency range where the hearing aid has a significant output power should be considered (i.e., in case of hearing aids E and H, frequencies from 650 Hz up to 5500 Hz). Thus, this measure is restricted to frequencies where output SNR is high.

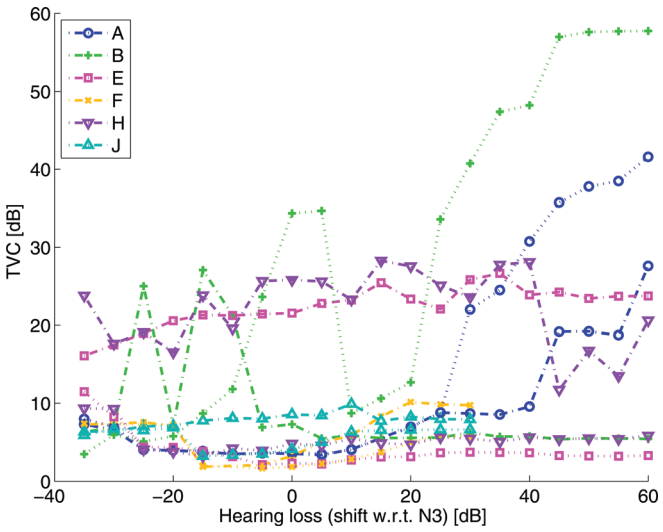

FIG. 5. Performance on the TVC measure for the organ stimulus. A value greater than 20 dB indicates instability. Although in the described implementation, TVC detects the largest peak or dip in the transfer function variation, the TVC due to a dip will rarely reach 20 dB.

2. PCR

The results using ΔPCR are depicted in Fig. 6(a) and those obtained using PCR_{norm} are presented in Fig. 6(b). Unlike TVC, PCR only considers peaks, so it is not affected by adaptive gain reduction such as notch filtering.

Compared to the TVC in Fig. 5, weak oscillations are less pronounced. Also, from the definition of ΔPCR, the measure is directly proportional to the narrowband signal energy. Thus, high energy oscillation frequencies are weighted more. As the measure takes the power at the oscillation frequencies into account, it is less susceptible to noise and estimation

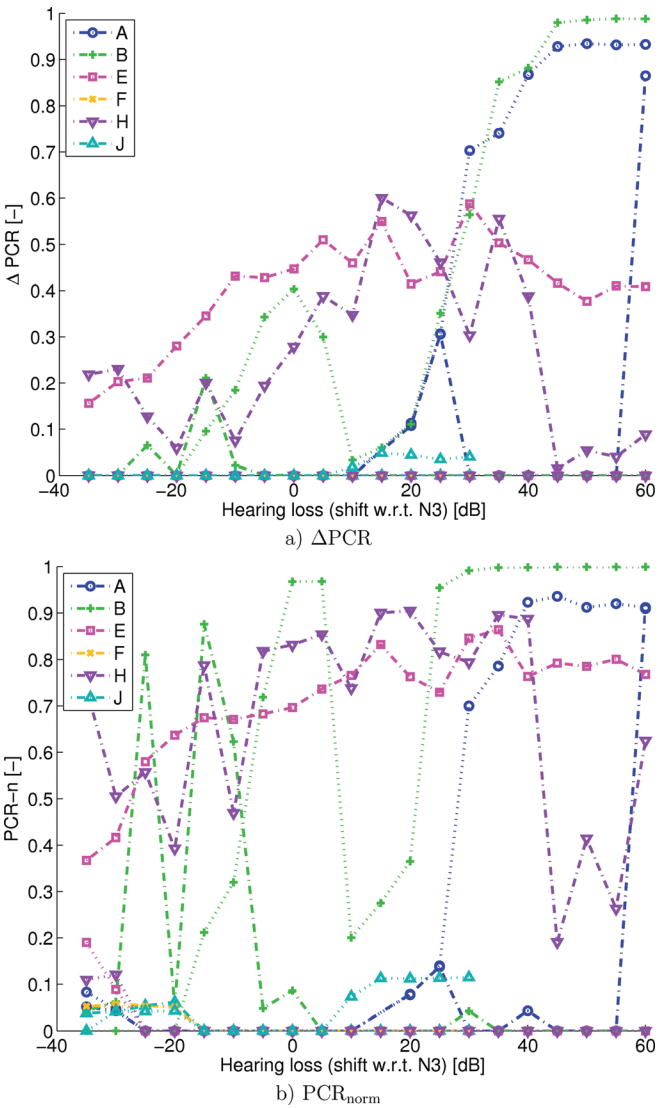

FIG. 6. Performance on the PCR measures for the organ stimulus. Values over 0.5 indicate instability (Ref. 14).

Downloaded 12 Mar 2012 to 134.58.253.57. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp

errors. However, due to this energy dependent weighting, oscillations at frequencies with low energy are not easily detected (e.g., device B for the organ stimulus for hearing losses of − 25 and − 15 dB.)

Results using PCR_n are similar to TVC. However, in contrast to ΔPCR, the power at the oscillation frequency is not taken into account. Thus weak oscillations may also result in large values, allowing for the detection of low energetic high frequencies (e.g., device B). However, similar to the TVC in this case, such independence from the signal power makes this measure susceptible to noise and estimation errors.

C. Measures for detecting signal distortions

1. SD

Figure 7 shows the performance of the devices when using the log spectral distortion to quantify the effect of feedback. Similar to FSR, the distortion gradually increases with increasing hearing loss.

With reference to Fig. 7, we also see that the distortion when feedback control is activated is *higher* for devices E, F, H, and J than when it is disabled, indicating, again, that the respective feedback control systems have problems with tonal signals. However, unlike FSR, the spectral distortion measure is unaffected by phase modifications in the forward path.

Strong feedback results in large spectral distortion values, thus detection of instability can be done very well. However, results of this measure should be treated with some caution as gain modifications in the signal processing path may also contribute to SD (e.g., notch filtering in device A, gain reduction and noise suppression in device J).

As in the case of $FSR_{intellig}$, feedback at higher frequencies will be weighted less (at higher frequencies the weight factor is spread out over an increasingly larger bandwidth. Consequently, the weight allocated to each narrowband frequency reduces). As a result, the measure generally increases less with increasing feedback than the non-intelligibility weighted variants, and weak feedback at higher frequencies may not be well detected.

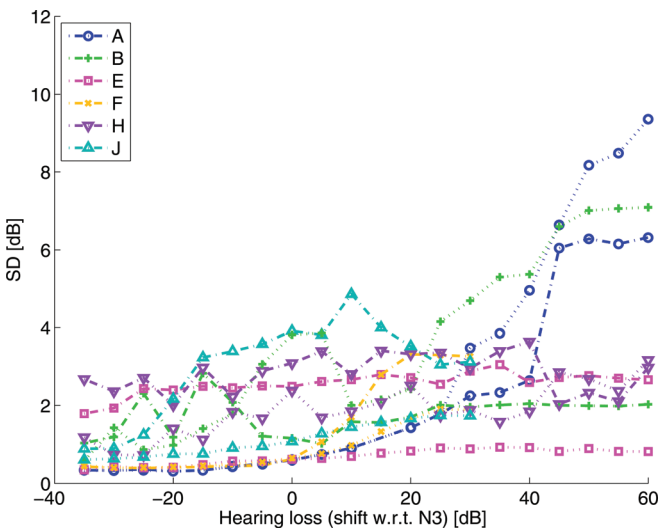

FIG. 7. Performance on the signal distortion measure for the organ stimulus.

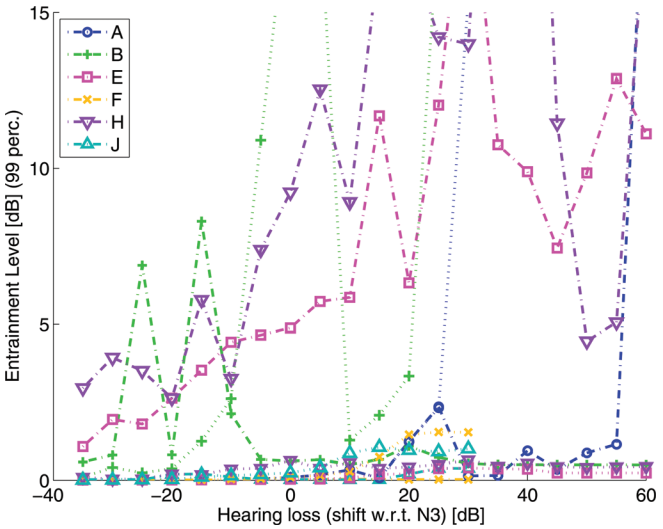

FIG. 8. Entrainment level (computed as the 99 percentile value over the complete recording) for the organ stimulus. A value over 5 dB is considered annoying.

While this measure indicates distortion in the output, properties of the auditory system (e.g., auditory masking) are not taken into account when computing the SD. Therefore, the perceptual relevance of the measure is not certain.

D. Entrainment level $\mathcal{L}_{\mathcal{H}}$

The entrainment measure is sensitive to artifacts in the signal. Thus, in addition to measuring instability, it can also be used to measure the distortions due to the feedback control system (e.g., bias for tonal signals). This is presented in Fig. 8. Instability and oscillations are manifested by large values of the entrainment level (devices B, E, and H). Results are similar to that of TVC and PCR_n.

As the measure is perceptually motivated, it may have a strong perceptual correlate and a comparison between FBC-on and FBC-off could yield a (perceptual) performance index for the feedback control system.

However, as the measure computation is based on a comparison with the short-term PSD of reference signal, it may be biased by gain modifications in signal processing path. Also, distinction between distortions due to the feedback and distortions caused by other processing in the hearing aids is difficult.

E. Gain measures

1. PSD

The PSD for the audiogram corresponding to a shift of 45 dB with respect to the standard N3 audiogram is presented in Fig. 9, for device A, a white noise input signal, and for the "no obstruction" condition. Feedback occurs here when feedback control is deactivated.

This measure offers the most information about the underlying processing. For the case with the feedback system disabled (dashed curve), the peaks corresponding to feedback are clearly visible (compare with the reference signal (solid curve)). For the system with the feedback system

Downloaded 12 Mar 2012 to 134.58.253.57. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp

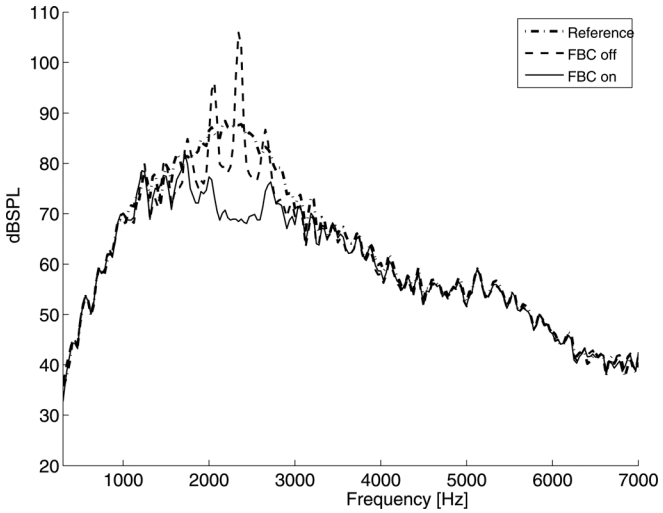

FIG. 9. PSD of the hearing aid output signal with feedback control disabled and enabled, and the reference signal for device A.

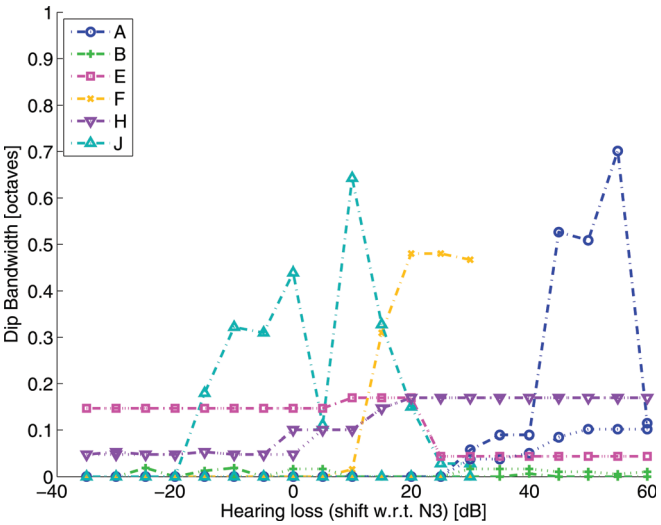

FIG. 10. Performance on the dip bandwidth measure for the organ stimulus.

enabled, reduction in feedback is evident. Also evident is the presence of adaptive gain reduction system, in this case a notch filter around 2 kHz.

This measure gives information on actual output level at all frequencies and hence, is useful for determining the useful frequency range for noise-sensitive measures (e.g., TVC). A comparison with the PSD of the reference signal can indicate feedback (manifested as peaks) and also adaptive gain reduction (e.g., notch filtering). Comparison of the output PSD for increasing hearing losses can also indicate static gain limitation by the hearing aid. Performing a similar comparison of the output PSD for different input signal levels can indicate the presence of compression. As the measure has a good frequency resolution, detection of narrow band oscillation frequencies is also possible.

The analysis of results obtained on this measure is time-intensive as multiple curves for different input levels need to be compared and it is not straightforward to condense the output of this measure into a single figure of merit.

2. DBW

Dip bandwith based system quantification is shown in Fig. 10. Large dip bandwidth is observed for devices A, F, and J. Device A has notch filtering as part of the feedback control system, whereas devices F and J have adaptive gain control and noise reduction algorithms active. This may explain the large values for these devices (cf. large signal distortion valus for A, F, and J).

Dip bandwidth provides a single figure of merit on output signal fidelity as compared to the PSD. It can also indicate gain reduction in the output. It is principally useful to better understand results of other measures (e.g., FSR and ORR), e.g., it can allow us to discriminate between distortion due to gain reduction and due to feedback.

F. Reproducibility

To analyze the reproducibility of the analysis, test and retest measurements were made for all the conditions and stimuli for devices A, B, and H. Reproducibility is indicated here on the basis of the hearing aid output signals for the test and retest. If a good agreement is obtained here, it might be safe to assume that the measure values are also reproducible.

Denoting the hearing aid outputs for the test measurement and the retest measurement by $u^1[k]$ and $u^2[k]$, the reproducibility was measured as

$$\gamma = \frac{E\left\{\left(u^{(1)}[k] - u^{(2)}[k]\right)^2\right\}}{\sqrt{E\left\{\left(u^{(1)}[k]\right)^2 E\left\{\left(u^{(2)}[k]\right)^2\right\}\right\}}}. \tag{19}$$

In general, the lower the value of γ, the better the reproducibility. To obtain a threshold for a "good" γ, we evaluate it for the hypothetical case when the difference in the $u^{(i)}[k]$ is just noticeable (JND) (i.e., we consider $u^1[k]$ to be a scaled version of $u^2[k]$, where the scaling is just perceivable).

If the scale values are taken such that the difference between the $u^{(i)}[k]$ lie in the range of 2–3 dB, it leads to γ values in the range $\gamma_{\text{thresh}} \approx [-18, -9]$ dB. These limits are subsequently taken as the threshold for quantifying the reproducibility. The γ values for all the stimuli and hearing losses are presented in the dB scale in Fig. 11.

Based on the above range of values for γ_{thresh}, it may be seen that devices A and H show a good reproducibility as long as the system is stable.[38] In the presence of instability, the reproducibility measure takes higher values, indicating lower correspondence between the outputs of the two runs. However, this is to be expected. Device B shows a stronger variation in its γ values because it is an RITC device and, for this device, the position of the receiver in the canal has a significant impact on the feedback path—especially close to instability. However, this positioning is not very reproducible and the effect of slight deviations is exaggerated when feedback occurs. Another aspect is the introduction of forward path modification by the feedback control system for this device—which may also not be reproducible across different measurements. This might account for the almost flat γ values for the case when feedback control is enabled. Note,

Downloaded 12 Mar 2012 to 134.58.253.57. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp

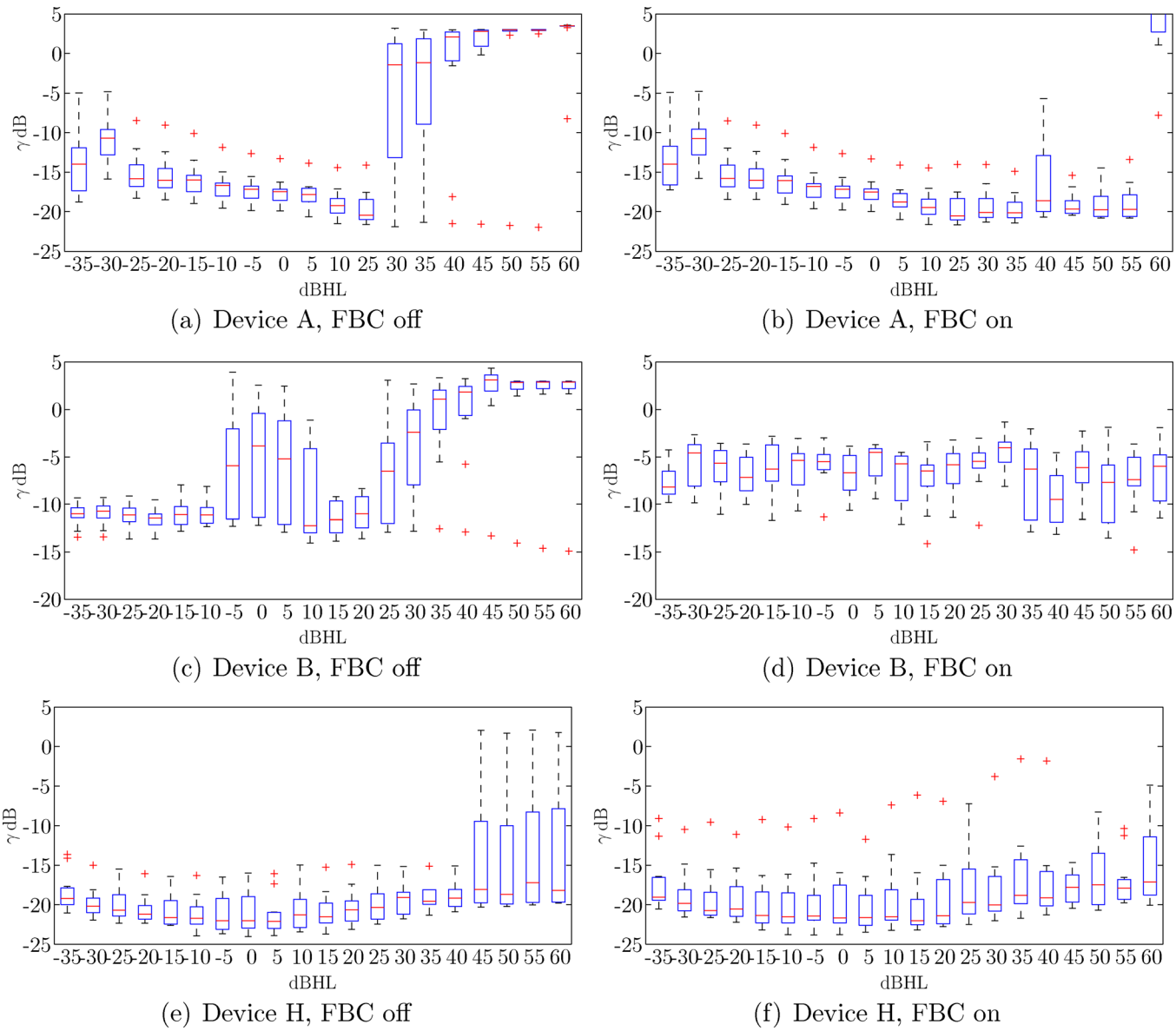

FIG. 11. (Color online) Reproducibility-γ of the hearing aid ouput over all stimuli and hearing losses for the selected devices. For stable conditions, devices A and H show good reproducibility in accordance with the JND thresholds defined in the text. Computing the measure on device B yields higher values of γ, both with and without the feedback system being enabled. The reasons for this are discussed in more detail in the corresponding section.

however, that this is due to the specific definition of γ, which might not be ideal for such devices.

There is another interesting by-product of the reproducibility measurement depicted in Fig. 11. Note the presence of the single outlier in the γ values for device A and H. Where the hearing aid is stable, this corresponds to the reproducibility measurement for the ISTS speech signal at 55 dB SPL, whereas when the devices are unstable, this corresponds to the impulsive sounds input. This result may be explained as follows: for the ISTS speech at 55 dB SPL presentation level, the signal to ambient noise ratio is low, leading to lower reproducibility. When the hearing loss is higher, feedback occurs. The impulsive noise signal, with long pauses between each sound, contains more feedback regions, thereby manifesting lower reproducibility.

G. Summary of applicability

In the previous sections we have seen that certain measures are better for characterizing certain aspects of feedback control system performance as compared to other measures. We summarize these findings in Table III, where the appropriateness of a measure for characterizing a certain aspect of feedback control system performance is indicated by a "+." No score indicates that the measure is not applicable to measuring that aspect.

Additionally, in the same table, the sensitivity of the measures to gain and phase modifications in the signal processing path are presented. Note that while all measures are affected somewhat by such modifications, the measures highlighted in the table are especially susceptible. Therefore, caution should be exercised when using these measures to quantify feedback control system performance when it is known that the forward path introduces gain and/or phase modifications. In this context, it is noted that compression and gain limitation also aid in feedback control. It is this aspect which makes it difficult for any one measure to fully quantify feedback control performance in black-box conditions.

Downloaded 12 Mar 2012 to 134.58.253.57. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp

TABLE III. Summary of the capabilities of different measures.

Feature	ORR	FSR	$FSR_{intellig}$	TVC	ΔPCR	PCR_{norm}	SD	Entrainment	PSD	DBW
Instability	+	+	+	+	+	+	+	+	+	
Weak feedback		+	+	+	+	+	+	+	+	
Oscillations and entrainment artefacts		+	+	+	+	+	+	+	+	
Adaptive gain limitation	+								+	+
Static gain limitation									+	
Phase modification in forward path		+	+							
Sensitivity to phase modifications		–	–							
Sensitivity to gain modifications				–			–			

Note, however, that the aim is not to explicitly quantify the feedback canceller performance of the hearing aid, but the feedback control performance of the overall system. Thus, for example, if the compressor can stabilize the system, the measures should not indicate instability.

The measures are computed with respect to a reference signal obtained using a closed fit at the same gain and compression settings, which is then transformed into an equivalent open-fit signal using the compensation filter $w[k]$. This reference signal is the HA output that would be perceived in the absence of acoustic feedback, including the effect of compressor (an ideal situation). Comparing this reference with the output with FBC on/off should principally demonstrate the effect of feedback on the full HA processing.

If there is acoustic feedback, the compressor behavior changes, and measures will be affected to some extent. But in such a case where compressor behavior changes to control feedback, the signal distortion would increase, which can be tracked using, e.g., the log-spectral distortion measure. Thus, the measures have to be used in conjunction with one another.

We also analyze the correlation *between* the results obtained from the different measures[39] using the Pearson correlation coefficient. This is presented in Table IV.

In all cases, except for the dip bandwidth (DBW) (last row and column in the table), the correlation is significantly non-zero. We further observe a good correlation between measures that are similar in concept (e.g., between the TVC and the PCR variants). We also note a significant non-zero correlation between DBW and TVC and DBW and SD, implying that a high value for TVC or SD might not necessarily be due to feedback, but could also be due to gain reduction—which is measured by the dip bandwidth. Such a cross-examination allows one to better distinguish between feedback presence and gain reduction. Thus, the dip bandwidth, which does not detect feedback *per se*, should be used more as an auxiliary tool to better analyze the results of the other measures. As a side note, the TVC may be made insensitive to gain reduction by considering only the peaks in the computation of this measure. We expect the correlation between this (modified) TVC and the DBW to then decrease.

We also see that the intelligibility weighted FSR measure correlates the least with the other measures. This may be explained as follows: feedback and harmonic distortions due to feedback have a narrowband nature and usually occur at high frequencies. However, intelligibility-weighted measures allocate weights to frequency *bands*, the bandwidth of which increases with frequency. Thus, the frequency-weighting is spread over a larger bandwidth at the higher frequencies allocating, as a result, lower weights to each narrowband component in this spectral range. Thereby, feedback distortions contribute less to the measure as compared to other approaches.

A closely related issue for the characterization of a feedback control system is also the proper choice of stimulus. Our experiments indicate that, all other conditions being the same, more feedback was obtained for the ISTS speech stimulus at low levels, for the impulsive sounds and for tonal signals. These represent, in our opinion, a more challenging scenario for any feedback control system. For the ISTS speech at low levels, the problem could be engendered by the higher compressor gains leading to larger amplification of weak sounds which, in turn, increase feedback. A similar reason may be attributed to the impulsive sounds, whereas tonal signals are highly correlated and affect convergence of the feedback cancellation algorithm or lead to target signal

TABLE IV. Pearsons correlation coefficient among the different measures.

	ORR	FSR	$FSR_{intellig}$	TVC	ΔPCR	PCR_{norm}	SD	Entrainment	DBW
ORR	–	0.820	0.427	0.745	0.824	0.704	0.654	0.89	– 0.023
FSR	0.820	–	0.666	0.771	0.809	0.69	0.768	0.756	0.0759 ($p < 0.05$)
$FSR_{intellig}$	0.427	0.666	–	0.528	0.394	0.480	0.433	0.359	0.086 ($p < 0.05$)
TVC	0.745	0.771	0.528	–	0.805	0.877	0.792	0.780	0.220 ($p < 0.05$)
ΔPCR	0.824	0.809	0.394	0.805	–	0.843	0.711	0.810	0.007
PCR_{norm}	0.704	0.690	0.480	0.877	0.843	–	0.553	0.723	– 0.020
SD	0.654	0.768	0.433	0.792	0.711	0.553	–	0.680	0.473 ($p < 0.05$)
Entrainment	0.890	0.756	0.359	0.780	0.810	0.723	0.680	–	0.031
DBW	– 0.023	0.076	0.086	0.220	0.007	– 0.020	0.473	0.031	–

Downloaded 12 Mar 2012 to 134.58.253.57. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp

cancellation—introducing distortions in the hearing aid output.

VII. CONCLUSIONS

This contribution has assessed the applicability of a range of instrumental measures for evaluating hearing aid feedback control system performance in black-box settings. In general, all measures can detect instability, except for dip bandwidth. However, detection of instability is not the purpose of the dip bandwidth. The power spectral density (PSD) offers the most information, however, it also requires time consuming analysis and does not offer a single figure of merit.

Intelligibility weighted measures (FSR and signal distortion) have lower values than the non-weighted counterparts. In such measures, the weights given to higher frequencies are lower than those given to lower frequencies. However, as feedback mainly occurs at higher frequencies, allocating a lower weight at these frequencies lowers the usability of these measures. Thus, intelligibility-weighted measures are not very applicable to feedback system evaluation.

All measures are biased by modification of the forward path gain by the feedback control system, adaptive gain reduction, notch filtering, etc. Therefore, during analysis, the measures should be used *in conjunction* with measures that can detect such modifications (e.g., dip bandwidth).

For detecting weak oscillations the FSR is especially suited. It is also the only measure that can detect phase modifications in the forward path.

Reproducibility tests indicate good reproducibility of the hearing aid output signals, below instability. The RITC device shows higher variations in the measures due to the difficulty in positioning the receiver in a reproducible manner.

In conclusion, this study has tested several devices from different manufacturers in black-box conditions, for realistic input signals and varying hearing loss. It has enabled us to see which measure is most suitable to examining a particular aspect of feedback control system performance, and serves as a beginning for conducting perceptual tests.

ACKNOWLEDGMENT

The authors would like to thank the hearing-aid companies participating in this research study. Specifically, we are grateful to Phonak, GNResound, Siemens, Oticon, and Widex for providing us with the hearing aids used in this study. We would also like to acknowledge the assistance of Ivo Merks of Starkey in implementing the entrainment measure. The work of N.M. is supported by the EU project AUDIS—Digital Signal Processing in Audiology. We thank Birger Kollmeier for constant support and fruitful discussions.

[1]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**, 147–161 (2003).

[2]J. E. Greenberg, P. M. Zurek, and M. Brantley, "Evaluation of feedback-reduction algorithms for hearing aids," J. Acoust. Soc. Am. **108**, 2366–2376 (2000).

[3]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**, 119–131 (2002).

[4]C. Boukis, D. P. Mandic, and G. Constantinides, "Toward bias minimization in acoustic feedback cancellation systems," J. Acoust. Soc. Am. **121**, 1529–1537 (2007).

[5]J. M. Kates, "Adaptive feedback cancellation in hearing aids," in *Adaptive Signal Processing: Applications to Real-World Problems*, edited by J. Benesty and Y. Huang (Springer-Verlag, Heidelberg, 2003), Chap. 2, pp. 23–55.

[6]J. A. Maxwell and P. M. Zurek, "Reducing acoustic feedback in hearing aids," IEEE Trans. Speech Audio Process. **3**, 304–313 (1995).

[7]M. G. Siqueira and A. Alwan, "Steady-state analysis of continuous adaptation in acoustic feedback reduction systems for hearing-aids," IEEE Trans. Speech, Audio Process. **8**, 443–453 (2000).

[8]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. **53**, 3749–3763 (2005).

[9]G. Grimm, V. Hohmann, and B. Kollmeier, "Increase and subjective evaluation of feedback stability in hearing aids by a binaural coherence-based noise reduction scheme," IEEE Trans. Audio Speech Language Process. **17**, 1408–1419 (2009).

[10]J. Hellgren, T. Lunner, and S. Arlinger, "Variations in the feedback of hearing aids," J. Acoust. Soc. Am. **106**, 2821–2833 (1999).

[11]J. M. Kates, "Room reverberation effects in hearing aid feedback cancellation," J. Acoust. Soc. Am. **109**, 367–378 (2001).

[12]B. Rafaely, M. Roccasalva-Firenze, and E. Payne, "Feedback path variability modeling for robust hearing aids," J. Acoust. Soc. Am. **107**, 2665–2673 (2000).

[13]M. R. Stinson and G. A. Daigle, "Effect of handset proximity on hearing aid feedback," J. Acoust. Soc. Am. **115**, 1147–1156 (2004).

[14]D. J. Freed and S. D. Soli, "An objective procedure for evaluation of adaptive antifeedback algorithms in hearing aids," Ear Hearing **27**, 382–398 (2006).

[15]D. J. Freed, "Adaptive feedback cancellation in hearing aids with clipping in the feedback path," J. Acoust. Soc. Am. **123**, 1618–1626 (2008).

[16]G. Grimm and V. Hohmann, "Combinations of monaural and binaural feedback control algorithms increase added stable gain," in *International Hearing Aid Research Conference (IH-CON)*, Presentation, Lake Tahoe, CA, 2006.

[17]S. Gao and S. Soli, "Method of measuring and preventing unstable feedback in hearing aids," US patent No. 6,134,329 (2000).

[18]I. Merks, S. Banerjee, and T. Trine, "Assessing the effectiveness of feedback cancellers in hearing aids," Hearing Rev. **13**, 53–57 (2006).

[19]M. Shin, S. Wang, R. A. Bentler, and S. He, "New feedback detection method for performance evaluation of hearing aids," J. Sound Vibration **302**, 350–360 (2007).

[20]A. Spriet, G. Rombouts, M. Moonen, and J. Wouters, "Adaptive feedback cancellation in hearing aids," J. Franklin Inst. **343**, 545–573 (2006).

[21]A. Spriet, M. Moonen, and J. Wouters, "Objective evaluation of feedback reduction techniques in hearing aids," in *Proceedings of the European Signal Processing Conference (EUSIPCO)*, Glasgow, UK, 2009, pp. 1860–1863.

[22]R. Huber and B. Kollmeier, "PEMO-Q. A new method for objective audio assessment using a model of auditory perception," IEEE Trans. Audio Speech Language Process. **14**, 1902–1911 (2006).

[23]T. Thiede, W. C. Treurniet, R. Bitto, C. Schmidmer, T. Sporer, J. G. Beerends, C. Colomes, M. Keyhl, G. Stoll, K. Brandenburg, and B. Feiten, "PESQ–The ITU-standard for objective measurement of perceived audio quality," J. Audio Eng. Soc. **48**, 3–29 (2000).

[24]J. Beerends, J. Krebber, R. Huber, K. Eneman, and H. Luts, "Speech quality measurements for the hearing impaired on the basis of PESQ," in *124th Audio Engineering Society (AES) Convention*, Amsterdam, The Netherlands, 2008, pp. 1–7.

[25]J. M. Kates and K. H. Arehart, "A speech quality metric based on a cochlear model," J. Acoust. Soc. Am. **125**, 2724 (2009).

[26]K. H. Arehart, J. M. Kates, M. C. Anderson, and L. O. Harve, "Effects of noise and distortion on speech quality judgments in normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **122**, pp. 1150–1164 (2007).

[27]K. Chung, "Challenges and recent developments in hearing aids. Part, I. I. Feedback and occlusion effect reduction strategies, laser shell manufacturing processes, and other signal processing techniques," Trends Amplif. **8**, 125–164 (2004).

Downloaded 12 Mar 2012 to 134.58.253.57. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp

[28]"Autoit v3," online resource http://www.autoitscript.com/ (Last accessed December 4, 2009).
[29]A. Spriet, M. Moonen, and J. Wouters, "Evaluation of feedback reduction techniques in hearing aids based on objective performance measures," J. Acoust. Soc. Am. **128**, 1245–1261 (2010).
[30]A. Spriet, K. Eneman, M. Moonen, and J. Wouters, "Objective measures for real-time evaluation of adaptive feedback cancellation algorithms in hearing aids," in *Proceedings of the European Signal Processing Conference (EUSIPCO)*, Lausanne, Switzerland, 2008, pp. 1–4.
[31]B. Moore, *An Introduction to the Psychology of Hearing*, 5th ed. (Academic Press, London, 2003), pp. 72–78.
[32]P. Vary and R. Martin, *Digital Speech Transmission: Enhancement, Coding and Error Concealment* (Wiley, New York, 2006), pp. 69–70.
[33]I. Merks, "Method to detect and quantify entrainment (bias) of a feedback canceller in a hearing aid," Technical Report, Signal processing group, Starkey labs, 2009.
[34]E. Zwicker and H. Fastl, *Psycho-acoustics. Facts and Models*, 2nd ed. (Springer Verlag, Heidelberg, 1999), pp. 223–226, 241–243.
[35]E. Zwicker, "Subdivision of the audible frequency range into critical bands (Frequenzgruppen)," J. Acoust. Soc. Am. **33**, 248 (1961).
[36]"2nd IEC/CD 60118-15 'Electroacoustics—Hearing aids—Methods for characterising signal processing in hearing aids,"' IEC Standards.
[37]N. Bisgaard, "Report on Selection of Standard Audiograms for the ISMADHA Measurement Procedure," EHIMA technical committee report, 2007.
[38]Device A has outliers for the two hearing losses at 15 and 20dB. Offline examination of the recorded signals indicate that the hearing aid was incorrectly programmed for these values of hearing loss. Therefore, these points are redacted in the presented figure.
[39]For this analysis, the data of device B have been excluded as the phase modifications introduced by the device in the forward path biases certain measures, as mentioned previously.
[40]M. Nilsson, S. D. Soli, and A. Sullivan, "Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise," J. Acoust. Soc. Am. **95**, 1085–1099 (1994).
[41]"EHIMA website," online resource http://www.ehima.com/ehima2/international_speech_test_signal_plus_terms_of_use_ists_v1.zip (Last accessed July 10, 2010).
[42]I. Holube, S. Fredelake, and M. Vlaming, "Entwicklung eines neuen internationalen Sprach-Testsignals (ISTS)" [Development of a new international speech test signal (ISTS)], Hörakustik, 2009, pp. 8–11
[43]M. Vlaming and I. Holube, "A new international speech test signal (ISTS) for the assessment of speech amplification in hearing aids," in *9th congress, European Federation of Audiological Societies (EFAS)*, abstract and presentation, 2009.

Downloaded 12 Mar 2012 to 134.58.253.57. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp