Objective estimation of loudness growth in hearing-impaired listeners

Ikaro Silva^{a)}

Laboratory for Computational Physiology, Harvard-MIT Division of Health Sciences and Technology,
Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139

Michael Epstein

Auditory Modeling and Processing Laboratory, Department of Speech-Language Pathology and Audiology,
Northeastern University, 360 Huntington Avenue, Boston, Massachusetts 02115

(Received 19 May 2011; revised 14 November 2011; accepted 15 November 2011)

A methodology for the estimation of individual loudness growth functions using tone-burst otoacoustic emissions (TBOAEs) and tone-burst auditory brainstem responses (TBABRs) was proposed by Silva and Epstein [J. Acoust. Soc. Am. **127**, 3629–3642 (2010)]. This work attempted to investigate the application of such technique to the more challenging cases of hearing-impaired listeners. The specific aims of this study were to (1) verify the accuracy of this technique with eight hearing-impaired listeners for 1- and 4-kHz tone-burst stimuli, (2) investigate the effect of residual noise levels from the TBABRs on the quality of the loudness growth estimation, and (3) provide a public dataset of physiological and psychoacoustical responses to a wide range of stimuli intensity. The results show that some of the physiological loudness growth estimates were within the mean-square-error range for standard psychoacoustical procedures, with closer agreement at 1 kHz. The median residual noise in the TBABRs was found to be related to the performance of the estimation, with some listeners showing strong improvements in the estimated loudness growth function when controlling for noise levels. This suggests that future studies using evoked potentials to estimate loudness growth should control for the estimated averaged residual noise levels of the TBABRs.

© 2012 Acoustical Society of America. [DOI: 10.1121/1.3666024]

PACS number(s): 43.64.Ri, 43.66.Cb, 43.64.Jb [BLM]

Pages: 353–362

I. INTRODUCTION

Listeners with hearing losses of primarily cochlear origin are known to exhibit a loudness growth function with a general shape that is neither a trivial function nor easily derived from their threshold (Marozeau and Florentine, 2007). In particular, it has been suggested that hearing-impaired listeners (HILs) can have loudness growth functions that exhibit one of two forms. Listeners may exhibit either classical loudness recruitment (for review, see Moore, 1998), where the slopes of the loudness functions for HILs are larger than normal at threshold values, but asymptotically reach normal slope values at higher levels, or softness imperceptions (Buus and Florentine, 2001), where the slopes of loudness growth functions for HILs are not significantly greater than normal at any sensation levels, but loudness at threshold is elevated, yielding an inability to perceive soft sounds. In the case of softness imperception, it is also implied that there is a decrease in the subjective “loudness” range. It is also possible that HILs have a loudness growth function that exhibits a mixture of both recruitment and softness imperception (e.g., a loudness growth function with an elevated loudness at threshold and a higher than normal slope at low sensation levels) (Marozeau and Florentine, 2007).

Given the wide range of types of loudness growth observed in individual HILs (Hellman and Meiselman, 1992), and that the type of loudness growth cannot be determined by thresholds and discomfort levels alone, there is interest in advanced procedures that could be used to estimate loudness growth for fully customized non-linear hearing-aid fitting. Furthermore, an objective procedure for estimating loudness growth could be a valuable tool in providing customizable hearing-aid fitting for those not capable of performing standard psychoacoustical tasks (such as in infants or cognitively impaired adults). A recent study (Silva and Epstein, 2010), investigated a series of objective methods for estimating loudness growth in normal listeners from tone-burst auditory brainstem responses (TBABRs) and tone-burst otoacoustic emissions (TBOAEs). While this method showed promising results in normal listeners, it was not clear if the performance of the loudness growth estimation technique would be accurate and robust enough to use for estimating loudness growth for a diverse group of HILs. In particular, studies looking into loudness growth and TBABRs have yielded mixed conclusions (Pratt and Sohmer, 1977; Wilson and Stelmack, 1982; Babkoff *et al.*, 1984; Davidson *et al.*, 1990; Serpanos *et al.*, 1997; Gallego *et al.*, 1999), the TBABR measurement variability and residual noise levels being a major source of contention. The Silva and Epstein (2010) method however, attempts to control for part of the variability in the recorded TBABR through the use of the Non-Stationary Fixed-Multiple-Point

^{a)}Author to whom correspondence should be addressed. Electronic mail: ikaro@mit.edu

(NS Fmp) statistic for measuring residual noise levels (see [Silva, 2009](#) for review). The estimated residual noise level is the expected amount of noise present in the waveform after averaging (calculated either from normal averaging or weighted averaging with both under non-stationary noise conditions). The NS Fmp is an adaptation of the original Fmp statistic ([Elberling and Don, 1984](#)), with the key modifications being the flexibility for allowing multiple noise sources of different power (following similar non-stationary analysis proposed by [Bendat, 1964](#)) and allowing for dynamically expandable window size in which the noise sources are considered locally stationary (for further comparison please see [Silva, 2009](#)). While previous research ([Don and Elberling, 1994, 1996](#)) has shown that measured evoked response amplitudes can be sensitive to residual noise levels, it is not clear how the new NS Fmp statistic (or its estimated residual noise level) in the TBABR is related to the TBABR performance in estimating the psychoacoustical loudness growth function. Investigating the relationship between the residual noise level and the loudness estimation performance will allow specific understanding of the extent to which residual noise levels can be a significant confounding factor. In addition, exploring such a relationship can also help establish a minimum acceptable residual noise level for adequate loudness estimation from TBABRs, facilitating comparison across studies and possibly reducing experiment time in cases of ideal recording conditions.

The specific objectives of this study are therefore to (1) verify the feasibility of using the objective loudness growth method described in [Silva and Epstein \(2010\)](#) for estimating loudness in HILs; (2) investigate the relationship between residual noise levels of TBABRs and the quality of loudness growth estimation, and (3) provide a public database accessible through PhysioNet ([Goldberger et al., 2000](#)), that will contain normal and HIL psychoacoustical and physiological responses to tone-burst stimuli across a wide range of levels.

II. METHODS

A. Listeners

Eight listeners (four females, four males, ages 40 to 85), with a known history of hearing impairment confirmed through clinical test (audiometric measurements made in the Northeastern University Speech-Language, and Hearing Center) and thresholds higher than 25 dB HL at least one octave frequency between 250 Hz to 8 kHz), participated in all measurements ([ANSI, 1996](#)). Listeners were tested on the ear that had the best (lowest) thresholds determined through a 2-interval, 2-alternative forced-choice procedure ([Levitt, 1971](#)) with the exact stimuli used throughout the experiment. This choice of ear was made to ensure that measurements could be made over the maximum physical and perceptual stimulus ranges. Three listeners were tested on the left ear and five listeners were tested on the right ear. Three listeners could not be tested at 4 kHz because their thresholds were higher than 85 dB peak-equivalent sound pressure level (peSPL). The subjects used in this experiment had the following 2-interval, 2-alternative forced-choice thresholds in dB peSPL for 1 kHz (4 kHz): N1 left ear 60, N2 right ear 60, N3 right ear 55 (60),

N4 left ear 60 (70), N5 right ear 35 (65), N6 right ear 50, N7 right ear 30 (35), and N8 left ear 55 (70).

B. Stimuli

The tone bursts used were 1-kHz tones with 4 ms duration and 4-kHz tones with 1 ms duration. The stimuli were multiplied with a Gaussian window and then end-padded with silence to generate a stimulus length of 41.7 ms. The Gaussian window was generated through the MATLAB function GAUSSWIN, which resulted in attenuations of about 27 dB at the start (0) and end times (1 or 4 ms) of the tone burst. The lack of a steady-state region in the window (i.e., no plateau) helped ensure consistent root mean square (RMS) values, equal spectral dispersion on a logarithmic scale, and a good TBABR response to transient stimuli. The stimulus levels varied from 5 dB below the listener's threshold to 100 dB peSPL in steps of 5 dB. Levels matched the specifications of the voltage-to-level conversion provided by Etymotic Research for the ER-10C apparatus.

For the cross-modality matching (CMM) and magnitude estimation (ME) procedures, each stimulus presentation consisted of 12 concatenated 41.7 ms intervals in order to generate a train of tone bursts that lasted approximately 0.5 seconds. This train was presented in place of a single tone-burst in order to compensate for any potential temporal-integration effects resulting from the continuous, rapid presentation of stimuli in the TBABR/TBOAE procedure ([Florentine et al., 1996](#); [Buus et al., 1997](#); [Zwicker and Fastl, 1999](#)). Levels were determined by a pressure-proportional voltage-to-level conversion based on calibration levels measured in a 6-cc coupler (B&K 4152, Nerum, Denmark). The stimuli used were identical to that used by [Silva and Epstein \(2010\)](#).

C. Apparatus

The stimuli were generated in MATLAB (2006b running on Ubuntu) and converted from digital (48-kHz sampling frequency) to analog using a 24-bit Lynx Two sound card. The analog signal was then passed through a Tucker-Davis Technologies HB7 headphone buffer and presented monaurally via the two transducers of the Etymotic ER-10C to a listener inside a double-walled sound-attenuating booth. During the TBOAE and TBABR measurements, the recordings from the ER-10C were converted from analog to digital (48 kHz sampling frequency) via a Lynx Two sound card. Routine calibration for each system was performed to test for proper wiring and ER-10C output in a plastic syringe coupler provided by Etymotic. All levels were determined using RMS of the windowed signal relative to the specifications which were provided by Etymotic and verified by doing an actual in ear measurement for a single listener using a Fonix 6500-CX real-ear system. The apparatus used was very similar to that of [Silva and Epstein \(2010\)](#).

D. Loudness growth estimation through CMM

Listeners were presented with six repetitions of each level in random order and asked to cut a string to be "as long as the sound is loud." After the listener cut each string, they taped it into a notebook and turned the page. Two blocks of trials were run separately, one for each of the two test

frequencies. The stimuli consisted of tone bursts from 5 dB below threshold to 100 peSPL varying in 5 dB steps presented in a random order. If a particular stimulus was not heard, no string was cut. A secondary set of threshold values were estimated from the CMM in three steps. First, a psychometric function was estimated by measuring the percentage of trials for each peSPL for which the listeners provided a string (i.e., they heard a sound). After the first step, a first-order polynomial was fitted between the data-point at the lowest peSPL level that had a 100% response rate and the data-point at the highest peSPL that had a 0% response rate. The threshold value was finally estimated by the peSPL value of the fitted first-order polynomial that yielded an estimated 50% response on the psychometric curve. This second set of thresholds (averaged with a similar threshold obtained from the ME described below) was obtained in order to determine the starting level of the stimuli for the TBABR and TBOAE recordings.

The loudness estimate for each level was the transformed geometric mean of the string lengths produced for that level. The transformation was performed in response to the finding that CMM, though it provides access to the details of the shape of the loudness function for individual listeners, yields functions with shallower slopes than other procedures (Epstein and Florentine, 2005). As such, a single string-length multiplicative correction factor, the same as that used in Silva and Epstein (2010), was used to correct the shallower slopes resulting from this procedure. This correction factor was previously determined for the procedure using normal-hearing data and was selected to match the average group data to a power function with an exponent equal to 0.3. This power function was chosen because of its simple approximation to the general form of the loudness function (Stevens, 1955, 1957, 1961; Hellman and Zwislocki, 1963; Stevens and Guirao, 1964). This constant, previously used correction factor was then applied to the individual data. The final loudness growth curve was subtracted by an offset in order to yield a zero-mean loudness curve for comparison with loudness curves obtained through other modalities.

Estimates for a fitted power-function exponent, (i.e., slope) of the loudness growth function derived from the CMM data were obtained by fitting to a linear, least-squares best fit to the CMM data points. The slope of the fitted line was multiplied by a factor of 10 in order to yield slope values in terms of the exponent of a power function, which is a common way of describing loudness slope (Stevens, 1955). Two other loudness growth estimates were also obtained from the final CMM procedure by using polynomial and INEX fitting (described later). These two smoothed estimates were done in order to provide additional references for the estimation procedures that utilize similar noise control methods.

E. Loudness growth estimation through ME

Listeners were presented with a series of tone bursts and asked to enter a numerical value that corresponded to the loudness of the stimulus. The listeners were told to use any positive number to enter 0 if no sound was heard. They were also encouraged to use decimals. The stimuli consisted of

tone bursts from 5 dB below threshold to 100 peSPL varying in 5 dB steps presented in a random order. Prior to the measurements, there was a practice test in which each level was presented only once in order to get the listener acquainted with the overall stimuli range. Each of the stimuli presented eight times in random order. If a listener entered 0 more than four times for a given level, that level was not used (treated as sub-threshold) during data analysis. The final estimate for a specific level was calculated from the geometric mean of the non-zero numbers. Threshold values estimated from the ME data were calculated by fitting a polynomial to the percent of non-zero entries as a function of level and interpolating the peSPL at which the polynomial value was 50%. Similar to the CMM method, two additional loudness growth estimates were obtained from the ME procedure by using polynomial and INEX fitting, described later. Loudness-growth power-function slope estimates were obtained in the same manner as those for the CMM data.

F. TBOAE and TBABR recordings

The TBOAEs and TBABRs were recorded simultaneously in a sound-attenuating, electrically shielded booth. Listeners had three electrodes affixed: the non-inverting electrode was positioned on the forehead, the inverting electrode was positioned on the ipsi-lateral mastoid (behind the ear), and the ground electrode positioned on the contralateral mastoid. Listeners were scrubbed with alcohol at the locations prior to the electrode placement. The electrode signal was sent to a GRASS QP511 Quad AC Amplifier, where it was band-pass filtered from 30–3000 Hz, amplified by a factor of 50 000, and sent to a 24-bit Lynx Two Soundcard that recorded with a sampling rate of 48 kHz. Stimuli were presented in blocks of 1000 trials (41.7 ms per trial), at a presentation rate of about 24 Hz. Each block of trials was repeated at least four times for each level yielding at least 4000 recordings per level in a session. All listeners participated in at least 3 sessions on separate days, yielding a total of at least 12 000 trials per level for each frequency. More trials were collected for some listeners that were available to participate in additional sessions (the complete data for every trial is available at www.physionet.org). Listeners L3 and L6 participated in one additional session to collect a total of 16 000 trials each, and listener L5 participated in a total of eight sessions, yielding 24 000 trials for each level. For each level, two averages of TBOAE recordings were made. The first average consisted of a weighted mean (Elberling and Wahlgreen, 1985) of all the trials in the first half of each of the 12 blocks (total of at least 6000 trials). The second average consisted of a weighted mean of all the trials in the second half of each of the 12 blocks (total of at least 6000 trials). These averages were the basis for the loudness estimation procedure described in Epstein and Silva (2009). For each frequency, the stimuli were presented in ascending order from 5 dB below the listener's threshold to 100 dB peSPL in steps of 5 dB. For the TBABR recordings, an artifact rejection threshold of $50 \mu V^2$ was applied. Throughout the experiment, a computer outside the booth displayed, for the current level, two weighted sub-averages

TABLE I. Descriptive statistics for the psychoacoustical measurements of loudness growth with respect to peSPL for 1-kHz tone bursts for HILs.

Lst	MSE (CMM vs ME)	CMM TH	ME TH	CMM Slope	ME Slope	CMM Stderr	ME Stderr
L1	0.014	70	70	0.41	0.48	0.09	0.06
L2	0.034	70	70	0.12	0.24	0.17	0.03
L3	0.258	60	60	0.27	0.51	0.07	0.12
L4	0.015	65	70	0.22	0.19	0.08	0.03
L5	0.035	45	45	0.22	0.13	0.07	0.03
L6	0.198	55	55	0.44	0.19	0.15	0.06
L7	0.008	40	40	0.14	0.17	0.05	0.03
L8	0.053	60	65	0.45	0.30	0.11	0.09
Ave	0.077	58	59	0.29	0.28	0.10	0.06
Median	0.034	60	62	0.25	0.21	0.09	0.05
Std	0.095	10	11	0.13	0.14	0.04	0.03

of the evoked response, the estimated SNR as a function of trial, the estimated noise variance as a function of trial, the estimated acoustical noise as a function of trial, and the estimated power of weighted average as a function of blocks of trials using the running NS FMP statistic (Silva, 2009). This information was used to monitor the recording and possible changes in the electrode placement and impedance. No session showed a significant and consistent increase in noise variance, and no patient reported removal or movement of either the electrodes or the insert earphone.

G. Estimation of loudness from TBOAEs

The procedure used to estimate the loudness growth from TBOAEs was exactly the same as that described by Epstein and Silva (2009), with the following parameters held fixed: window delay = 10 ms, window size = 20 ms, window type = Hanning, F-ratio = 2, where window delay determines the onset of the temporal analysis window after the stimulus onset, window size determines the length of the temporal analysis window, window type is the type of window that is applied to the data, and F-ratio is a parameter that determines the frequency bandwidth of the spectral analysis window. Briefly, the loudness estimation procedure for a single level was done by taking the cross-spectrum of two independent weighted average responses within the time and frequency ranges specified by the analysis parameters. The final loudness estimation for a particular level was made by summing all the positive real components of the estimated cross-spectrum.

Two additional loudness-growth estimates were obtained from the TBOAE procedure by using polynomial and INEX fitting. These two smoothed estimates were done

in order to provide additional references for the evoked potential estimation procedures that utilize the same noise control methods.

H. Estimation of loudness from TBABRs

The estimation of loudness growth from TBABRs was done in the exact same manner as that described in Silva and Epstein (2010). The estimation process consists of three stages: segmentation, a point-loudness estimate, and a noise constraint stage. The first stage is a segmentation stage where specific regions of the TBABR waveform are selected. A total of eight different segmentation techniques were used for segmenting the weighted averaged evoked response (Elberling and Wahlgreen, 1985; Silva, 2009): fullblock, abrblock, amlrblock, waveVamp, amlramp, fullsync, abrsync, and amlrsync. The first three segmentation techniques—fullblock, abrblock, and amlrblock—were based on the selection of waveform regions at fixed time intervals. The fullblock procedure selected the fixed region of the evoked response between 1 ms after the stimulus offset until the end of the trial (41.7 ms after the stimulus onset), the abrblock procedure selected a fixed region of the evoked response between 1 ms after the stimulus offset and 21 ms after the stimulus onset (the early component of the evoked response), and the amlrblock procedure selected a fixed region of the evoked response from 20 ms after the stimulus offset until 41.7 ms after the stimulus onset (later part of the response including the auditory middle latency response). The other three segmentation procedures: the fullsync, abrsync, and the amlrsync, were procedures whose time intervals were selected adaptively, but within the fixed region

TABLE II. Descriptive statistics for the psychoacoustical measurements of loudness growth with respect to peSPL for 4-kHz tone bursts for HILs.

Lst	MSE (CMM vs ME)	CMM TH	ME TH	CMM Slope	ME Slope	CMM Stderr	ME Stderr
L3	0.041	65	65	0.26	0.22	0.06	0.18
L4	0.005	80	75	0.20	0.14	0.06	0.02
L5	0.013	75	70	0.35	0.21	0.06	0.03
L7	0.001	40	45	0.16	0.16	0.05	0.05
L8	0.036	75	80	0.55	0.19	0.08	0.08
Ave	0.019	67	67	0.30	0.18	0.06	0.07
Median	0.013	75	70	0.26	0.19	0.06	0.05
Std	0.017	16	13	0.15	0.03	0.01	0.06

FIG. 1. Threshold vs log of perceptual range (difference in log between maximum and minimum responses). A significant linear correlation ($r = -0.69$, $p < 0.01$) was found between perceptual range and threshold, but variability within HILs still remains an important factor (i.e., see data points around 65 dB peSPL). The solid black line represents the normal average and the dotted black line represents -2 standard deviations from the average. Solid circles represent normal-hearing listeners and asterisks represent hearing-impaired listeners.

defined by their block counterpart described previously. The similarity analysis was done by comparing the current averaged waveform with a time aligned (zero-padded) version of the averaged waveform obtained at the next lower level. The similarity measure consisted of calculating the dot-product (i.e., the zero-lag cross-correlation) of the two waveforms in non-overlapping intervals of 1 ms. If the dot-product of the two waveforms was above an arbitrary threshold (set to 0), the 1 ms section was selected for further analysis, otherwise it was discarded. The final segmented waveforms consisted of these selected 1 ms regions. For additional details on all of the procedures, see [Silva and Epstein \(2010\)](#).

The second stage used all of the selected 1 ms regions from the first stage to generate a point estimate of the loudness at that particular level. The two point estimates used were the average RMS and the maximum amplitude squared. The maximum amplitude squared was chosen due to the use of Wave V amplitude for loudness estimation in previous literature ([Pratt and Sohmer, 1977](#); [Wilson and Stelmack, 1982](#); [Babkoff et al., 1984](#); [Davidson et al., 1990](#); [Serpanos et al., 1997](#); [Gallego et al., 1999](#)).

The third and final stage consisted of applying residual noise level constraints in a parametric fashion across loudness estimates obtained at all the levels. This stage utilized

residual noise values estimated from the weighted average waveforms ([Silva, 2009](#)), which are a measure of the “quality” of the TBABR, to control for the weight that each loudness point estimate has on the overall shape of the loudness growth function. Two procedures were used: weighted polynomial fitting and INEX fitting. The weighted polynomial fitting, a standard polynomial fitting procedure ([Scharf, 1991](#)), was used with the weights set equal to the inverse of the residual noise power of each level (note that this assumes that residual noise levels were uncorrelated across peSPL). The maximum order of the polynomial fitting was set to five, however the polynomial order was allowed to decrease if the data matrix was singular or badly scaled (determined from the rank of the fitting matrix). The second procedure, the INEX-fitting procedure, was based on the fifth-order polynomial model described first in [Florentine and Epstein \(2006\)](#) with the function parameters given in [Epstein and Silva \(2009\)](#); specifically $\log_{10}(N) = (1.7058 \times 10^{-9})L^5 - (6.587 \times 10^{-7})L^4 + (9.7515 \times 10^{-5})L^3 - (6.6964 \times 10^{-3})L^2 + (0.2367)L - 3.4831$, where N is the loudness in sones ([Stevens, 1936](#)) and L is the level in dB peSPL. The INEX is a loudness model based on temporal-integration and spectral-summation data ([Florentine et al., 1996](#); [Buus et al., 1997](#); [Buus et al., 1998](#)). The INEX fitting was done by using a family of horizontally shifted INEX curves and selecting the curve that minimizes the distance between the measured loudness point estimates weighted by their respective residual noise levels.

V. RESULTS

Tables I and II show the results for the psychoacoustical loudness-growth data obtained for eight HILs using 1-kHz tone bursts and four HILs using 4-kHz tone-bursts, respectively. The median estimated psychoacoustical threshold of the CMM (ME) for HILs was 60 (62) dB peSPL [these values for the [Silva and Epstein \(2010\)](#) study with normal listeners were 30 (35) dB peSPL]. Thus, two HILs had threshold values within a 15-dB range of the normal median: 45 (45) for L5 and 40 (40) for L7. Similar analysis for the 4-kHz data shows that the median CMM (ME) estimated thresholds were 75 (70) dB peSPL [these values for the [Silva and Epstein \(2010\)](#) study with normal listeners were 25 (30)

TABLE III. Statistics for the TBABR recordings on HI. The names represent V Lat—wave V latency; Ave Noise—average residual noise power (ηV^2) across peSPL; Std Noise—standard deviation of the residual noise power (ηV^2) across peSPL.

Lst	V Lat (1 kHz)	V Lat (4 kHz)	Ave Noise (1 kHz)	Std Noise (1 kHz)	Ave Noise (4 kHz)	Std Noise (4 kHz)
L1	7.1		0.2	0.1		
L2	9.4		1.1	1.0		
L3	7.5	4.7	1.1	0.9	1.0	3.7
L4	7.6	6.9	0.8	0.2	0.6	1.1
L5	7.7	4.6	1.4	0.2	2.0	4.6
L6	9.4		1.2	0.1		
L7	7.4	4.5	2.0	1.1	0.6	1.6
L8	7.4	6.3	0.6	0.3	0.8	4.0
Ave	7.9	5.4	1.0	0.4	1.0	3.0
Median	7.5	4.7	1.1	0.2	0.8	3.7
Std	0.9	1.1	0.5	0.4	0.5	1.5

FIG. 2. Median MSE across the HI. In general the performance of the wave V amp and fullsync are robust across frequency seem to perform the close to optimal within the procedures using evoked potentials. The values for the TBOAE procedure at 4 kHz was 0.41 (ME Ref) and 0.30 (CMM Ref).

dB peSPL]. Listener L7 had the lowest threshold 40 (45) of the all HILs, also within a 15 dB range of the normal median. The median fitted power-function slopes (i.e., Stevens' power function exponents) for HILs were 0.25 (0.21) at 1 kHz and 0.26 (0.19) at 4 kHz (within the range of, but slightly shallower than that of normal listeners from [Silva and Epstein \(2010\)](#)), with slopes for normal listeners of 0.31 (0.29) at 1 kHz and 0.3 (0.27) at 4 kHz.

Figure 1 shows a plot of the threshold value versus the log of the averaged CMM and ME perceptual range for all listeners [including data from the normal listeners obtained from [Silva and Epstein \(2010\)](#)]. The perceptual range was calculated by subtracting the log of minimum loudness value (the average number or string length a listener entered for this threshold level) from the log of the maximum loudness value (the largest number or string length a listener entered at 100 dB peSPL). The final range for each listener was the average range between the CMM and ME procedures. It is clear that listeners with the same threshold can have significantly different loudness perceptual ranges under the same set of stimuli. For instance, the threshold values around 65 dB peSPL seem to result in large individual

differences in perceptual ranges, with some listeners (exhibiting recruitment-type loudness growth) within two standard deviations of the normal-hearing listeners, and some listeners (exhibiting softness imperception-type loudness growth), with ranges of more than two standard deviations below the normal range. A significant ($p < 0.01$) linear correlation of $r = -0.697$ was found between the log perceptual range and threshold.

A. TBABR and TBOAE loudness growth estimation with no noise control

Table III shows descriptive statistics for the evoked response. The listener with highest residual noise levels for both frequencies was listener L5 and the listeners with the lowest residual noise levels for were L1 at 1 kHz, and L4 and L7 at 4 kHz. The median noise levels for 1 and 4 kHz were 1.1 and 0.8, respectively. The latency values for wave V represent averages across all noise levels using the algorithm described in Sec. II. It is important to highlight that latency values for wave V amp were estimated from the objective procedure wave V amp described in the Methods

FIG. 3. Individual loudness growth estimates for HILs using 1 kHz tone-burst.

FIG. 4. Individual loudness growth estimates for HILs using 4 kHz tone-Burst.

section, and due to its simple nature, they are not as accurate as a human annotator (a detailed example of the procedure is shown in [Silva and Epstein, 2010](#)). In fact, it's possible that this method might yield latency values smaller than normal due to spurious noise or artifact in the average waveform as suggested by values less than 5.5 ms. Future improvements on this method will likely require some smoothing across levels. As the goal of the present work was to only use automated procedures, these procedures may not be appropriate for latency estimation in general.

Figure 2 shows the loudness growth estimation performance, in terms of the median MSE, calculated across all the HILs. In general, the estimation done by the TBOAE was poorer compared to the estimations done with evoked potentials (e.g., wave V amp, and fullsync). The performances of the estimators that utilize evoked responses are not as close as the psychoacoustical estimators (particularly at 4 kHz), but are still well below the upper bound of the 4 kHz TBOAE which was 0.41 (ME Ref) and 0.30 (CMM Ref). The 4 kHz TBOAE is taken as an upper bound on the performance because of its equivalence to a trivial estimation linearly related to the stimulus intensity [probably a result of artifact contamination as observed in [Epstein and Silva \(2009\)](#) and [Silva and Epstein \(2010\)](#)]. Therefore, listeners who exhibit linear loudness growth (i.e., those with no compression and no recruitment) might match this response by coincidence.

Figures 3 and 4 show the individual loudness estimates obtained with the full sync procedure along with the TBOAE, CMM, and ME data.

B. TBABR loudness growth estimation with noise control and parametric fitting

Table IV compares the results in terms of the median MSE calculated across the HILs (HIL'S) for loudness growth estimation obtained using the two different noise control methods. For the psychoacoustical procedures and TBOAE estimators, the noise control methods were applied with the residual weights all set to 1 (uniform) across all peSPLs. The TBOAE estimation yields higher MSE compared to the other procedures. The wave V amp procedure also does not show a drastic drop in noise control as seen with normal listeners. In general the evoked potential estimators almost show a monotonic improvement in performance with respect to the

three different conditions, achieving equivalent performance with the psychoacoustical reference (except perhaps at the 1 kHz ME Ref condition).

Figures 5 and 6 show the individual loudness growth estimation for the fullsync procedure using the INEX noise control method along with the CMM and ME loudness growth estimates (Figs. 3 and 4 show the estimates but with no noise control applied). In general it seems like listeners L3, L5, and L8 yielded the best results across the different conditions (the estimation improvement in L5 at 4 kHz is particularly striking).

C. TBABR loudness growth estimation performance and residual noise levels

In order to investigate the effects of TBABR post-average residual noise levels on the performance the estimation of loudness growth function, two analyses of the data were performed. One analysis investigated the averaged evoked response residual noise level in all listeners versus the MSE and linear correlation coefficient between the TBABR wave V amp estimator and the CMM psychoacoustical

TABLE IV. Median MSE calculated across the HILs for loudness growth estimation obtained using the two different noise control methods (Wpoly and INEX). For the psychoacoustical procedures and TBOAE estimators, the noise control methods were applied with the residual weights all set to 1 across all peSPLs.

	Psychoacoustic	OAE	Wave V Amp	Full Block	Full Sync
1 kHz CMM Ref					
No Control	0.034	0.188	0.058	0.069	0.073
Wpoly	0.03	0.166	0.053	0.067	0.057
INEX	0.03	0.123	0.029	0.029	0.027
1 kHz ME Ref					
No Control	0.034	0.158	0.056	0.074	0.074
Wpoly	0.024	0.157	0.034	0.045	0.06
INEX	0.031	0.081	0.034	0.042	0.048
4 kHz CMM Ref					
No Control	0.013	0.308	0.044	0.071	0.064
Wpoly	0.012	0.317	0.04	0.038	0.03
INEX	0.013	0.337	0.023	0.013	0.015
4 kHz ME Ref					
No Control	0.013	0.411	0.029	0.084	0.076
Wpoly	0.013	0.423	0.0255	0.052	0.077
INEX	0.039	0.436	0.037	0.037	0.036

FIG. 5. Individual loudness growth estimation for HILs using 1 kHz tone-burst and the CMM, ME and fullsync procedures with INEX noise control applied.

estimator. The second investigation looked into how the MSE and linear correlation coefficient between CMM and the best TBABR estimated loudness growth functions varied as function of residual noise level when the number of trials of the averaged TBABR waveforms varied from 1000 to 12000. This later analysis used the data from the listener that yielded the lowest MSE between the TBABR and CMM estimates (L5 at 1 kHz).

Figure 7 shows the results for the first analysis. The top graph shows the estimated averaged evoked response residual noise level vs the MSE between wave V amp estimator and the CMM psychoacoustical estimator (correlation coefficient of $r=0.35$ with $p=0.05$). Figure 7 (bottom) shows a plot of the estimated averaged evoked response residual noise level and the linear correlation coefficient between the wave V amp estimator and the CMM estimator. In this case, there is a significant negative linear relationship between averaged evoked response residual noise level and a correlation between the two modalities ($r=-0.38$, $p=0.03$). If a MSE value of 0.15 is the cutoff for a reasonable fit based on the performance seen from Figs. 3–6, then it seems that setting an average value of WNS Fmp to be between 0.1 and 0.15 ηV^2 will yield, on average, a 0.15 MSE using wave V amp. The linear correlation coefficient plot shows a similar behavior, but with wider scatter; in this case setting the WNS Fmp to less than or equal to 0.0001 will yield on average a correlation coefficient of 0.7 between the two loudness growth estimated from the V amp procedure and CMM.

Figure 8 shows the results for the analysis on a single listener where the number of trials in the averaged physiological waveform was varied from 1000 to 12000 trials per level (in steps of 1000 trials). This yielded 12 data points with different averaged WNS Fmp values. No noise control method was applied to the evoked response loudness growth estimators. The mean residual noise level of the weighted averaged evoked response waveform was monotonic with respect to the number of trials used (i.e., the point to far left corresponds to using all 12000 trials and the point to the far right corresponds to using the first 1000 trials). The top graph shows the estimated averaged evoked response residual noise level vs the MSE (calculated between the evoked response estimators and the CMM psychoacoustical estimate). The estimated linear correlation for the top panel was $r=0.9612$ ($p<0.001$) for the full sync procedure and $r=0.7748$ ($p<0.001$) for the wave V amp procedure. Figure 8 (bottom) shows a plot of the estimated averaged evoked response residual noise level and the linear correlation coefficient between two evoked response loudness growth estimators and the CMM loudness growth estimator (wave V amp as stars, and full sync as filled triangles). The estimated linear correlation for Fig. 8 (bottom) was $r=-0.9697$ ($p<0.001$) for the full sync procedure and $r=-0.9844$ ($p<0.001$) for the wave V amp procedure. Overall, Fig. 8 shows that both the MSE and correlation coefficients of the wave V amp procedure are consistently worse than the full sync procedure. The graph that describes

FIG. 6. Individual loudness growth estimation for HILs using 4 kHz tone-burst and the CMM, ME and fullsync procedures with INEX noise control applied.

FIG. 7. Performance of the estimators as function of WNS Fmp for all listeners. (top) The MSE and (bottom) the linear correlation between the wave V amp procedure and CMM.

the correlation between the psychoacoustical estimate and the evoked response estimates [Fig. 8 (bottom)] seems to show ceiling effects for the wave V amp with the ceiling being at about 0.8, reached at about an average evoked response residual noise level of 0.2 (and performance slightly higher than 0.95). This ceiling behavior is also somewhat apparent in the MSE data as a floor effect, with the wave V amp and full sync procedures converging on MSE values of around 0.04 and 0.01 for WNS Fmp lower than 0.2. This suggests that the final performance, using 12 000 trials, was not that different from that using half the number of trials.

VI. DISCUSSION AND CONCLUSION

An analysis of the perceptual range for the all listeners [including normals from [Silva and Epstein \(2010\)](#)] as a function of their threshold showed that there could be large variability in perceptual range between two listeners with the same threshold (Fig. 1). This supports the consideration of at least two different existing models of loudness growth (softness imperception and recruitment) and is in qualitative agreement with the observation that the standard deviation of estimated slopes of the loudness growth function for a group of people increase as a function of the average hearing

threshold of the group ([Hellman and Meiselman, 1992](#)). From a more practical point of view, this variability underscores the general idea that simply knowing a listener's threshold is not sufficient to determine the person's perception of loudness as a function of stimulus level. Procedures that use categorical scaling of loudness, which typically use only seven unique categories, and measurements of discomfort level, might mask this perceptual range deficit because a perceptual range scale is already being imposed on the listener. Recently [Al-Salim et al. \(2010\)](#) have achieved very promising results with categorical loudness scaling, showing that consistent results for estimating perceptual range and appropriate gain can be obtained when using two linear segments to fit the loudness growth functions as oppose to one ([Elberling, 1999](#)). Further investigation is still needed, however, in order to determine if a categorical loudness scale induces a bias by imposing limits on the set and a growth relation perceived either by the ratio of the bars displaying the categories, or the difference in the numbers associated with them. Perhaps a study that looks into categorical loudness scaling, CMM, ME, TBABR, and TBOAE in the context of both correlation and MSE can help elucidate any biases among these modalities. [Epstein and Silva \(2009\)](#) suggested that the MSE is a better measure of fit for loudness functions than the Pearson correlation and no such analysis comparing all of these methods exists in the literature.

In terms of the overall performance of the estimators, the full sync procedure with the INEX noise control was the best estimate from the evoked response, and, although TBOAE performance at 1 kHz was reasonable, the full sync ABR procedure's performance was superior under all conditions (i.e., Table IV). It is possible that part of the decreased performance of the TBOAE procedure could be due to the parameter selection based only on normal young adults, ages 19–31 years (see [Epstein and Silva, 2009](#), for details). A parameter selection controlling for age might have improved the TBOAE results with the HI, since on average they constituted an older group (TBOAEs are known to yield different responses according to age, for review see [Hall, 2006](#)). The WNS Fmp residual error analysis shows a correlation between residual noise levels and the performance of the estimators. This corroborates the importance of controlling for residual noise levels when attempting to compare the amplitude of evoked responses across levels. Perhaps part of the controversy in the scientific literature regarding the relationship between ABR amplitude and loudness (see [Silva and Epstein, 2010](#), for a review) can be explained in terms of the relatively small number of trials collected in some studies (where no relationship was found across the estimated loudness growth functions), and the lack of specific assessment or estimation of the residual noise levels in the averaged physiological waveforms. For instance, an obvious improvement in the estimation of the loudness growth for listener L5 at 4 kHz was obtained by controlling for residual noise levels using the INEX method (Figs. 4 and 6). Thus, this study also supports the conclusion of previous works by [Don and Elberling \(1994, 1996\)](#), which highlighted the weakness of submitting all listeners to the exact same number of trials. Number of trials is not a sufficient condition to

FIG. 8. WNS Fmp level vs Performance for two TBABR procedures, wave V amp, and full sync, measured against the CMM psychoacoustical procedure for listener L5 as the number of averaged trials in the TBABR varied (see text for more details).

guarantee equal quality of waveforms across stimulus level, listeners, and experiments.

The current work also attempts to investigate the relationship between the performance of the evoked response loudness growth estimator and the estimated averaged evoked response residual noise levels. Using data from the “best” listener, a liberal strategy suggests that WNS Fmp values roughly around 0.2 generate appropriate conditions for loudness growth estimation from the auditory evoked responses. On the other hand, a conservative strategy based on data from all the listeners, suggests an average lower bound of the WNS Fmp between 0.1 and 0.15. The data does not suggest, however, that a low average WNS Fmp is a necessary condition for accurate loudness growth estimation, since on a very small number of occasions, good performance was obtained at relatively high WNS Fmp values (which may also help explain why some of the studies on Table I found a relationship with ABR and loudness using so few trials).

ACKNOWLEDGMENTS

The authors would like to thank two reviewers as well as the following people who helped with data collection: Mary Beth Bramel, Guinevere Deevy, Nora Rosenfeld, and Shoshannah Kantor. The authors would also like to thank Anna Neumann for help in preparing the final manuscript. This work was supported by NIH NIDCD 1R03DC009071.

- Al-Salim, S. C., Kopun, J. G., Neely, S. T., Jesteadt, W., Stiegemann, B., and Gorga, M. P. (2010). “Reliability of categorical loudness scaling and its relation to threshold,” *Ear Hear.* **31**(4), 567–578.
- ANSI. (1996). *American National Standard Specification for Audiometers* (American National Standards Institute, New York).
- Babkoff, H., Pratt, H., and Kempinski, D. (1984). “Auditory brainstem evoked potential latency-intensity functions: A corrective algorithm,” *Hear. Res.* **16**, 243–249.
- Bendat, J. S. (1964). “Mathematical analysis of average response values for non-stationary data,” *IEEE Trans. Biomed. Eng.* **11**(3), 72–81.
- Buus, S., and Florentine, M. (2001). “Growth of loudness in listeners with cochlear hearing losses: Recruitment reconsidered,” *J. Assoc. Res. Otolaryngol.* **3**, 120–139.
- Buus, S., Florentine, M., and Poulsen, T. (1997). “Temporal integration of loudness, loudness discrimination, and the form of the loudness function,” *J. Acoust. Soc. Am.* **101**, 669–680.
- Buus, S., Müssch, H., and Florentine, M. (1998). “On loudness at threshold,” *J. Acoust. Soc. Am.* **104**, 399–410.
- Davidson, S. A., Wall, L. G., and Goodman, C. M. (1990). “Preliminary studies on the use of an ABR amplitude projection procedure for hearing aid selection,” *Ear Hear.* **11**, 332–339.
- Don, M., and Elberling, C. (1994). “Evaluating residual background noise in human auditory brain-stem responses,” *J. Acoust. Soc. Am.* **96**(5), 2746–2757.
- Don, M., and Elberling, C. (1996). “Use of quantitative measures of auditory brain-stem response peak amplitude and residual background noise in the decision to stop averaging,” *J. Acoust. Soc. Am.* **99**(1), 491–499.
- Elberling, C., and Don, M. (1984). “Quality estimation of averaged auditory brainstem responses,” *Scand. Audiol.* **13**(3), 187–197.
- Elberling, C. (1999). “Loudness scaling revisited,” *J. Am. Acad. Audiol.* **10**, 248–260.
- Elberling, C., and Wahlgreen, O. (1985). “Estimation of auditory brainstem response, ABR, by means of Bayesian inference,” *Scand. Audiol.* **14**, 89–96.
- Epstein, M., and Florentine, M. (2005). “A test of the Equal-Loudness-Ratio hypothesis using cross-modality matching functions,” *J. Acoust. Soc. Am.* **118**, 907–913.
- Epstein, M., and Silva, I. (2009). “Analysis of parameters for the estimation of loudness from tone-burst otoacoustic emissions,” *J. Acoust. Soc. Am.* **125**, 3855–3864.
- Florentine, M., Buus, S., and Poulsen, T. (1996). “Temporal integration of loudness as a function of level,” *J. Acoust. Soc. Am.* **99**, 1633–1644.
- Florentine, M., and Epstein, M. (2006). “To honor Stevens and repeal his law (for the auditory system),” in *Fechner Day*, edited by D. E. Kornbrot, R. M. Msetfi, and A. W. MacRae (International Society for Psychophysics, St. Albans, England), pp. 37–42.
- Gallego, S., Garnier, S., Micheyl, C., Truy, E., Morgon, A., and Collet, L. (1999). “Loudness growth functions and EABR characteristics in digisonic cochlear implantees,” *Acta Otolaryngol.* **119**, 234–238.
- Goldberger, A. L., Amaral, L. A. N., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., Mietus, J. E., Moody, G. B., Peng C.-K., and Stanley, H. E. (2000). “PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals,” *Circulation* **101**(23), e215–e220.
- Hall, J. W. (2006). *New Handbook for Auditory Evoked Responses* (Allyn and Bacon, MA), pp. 263–280.
- Hellman, R. P., and Meiselman, C. H. (1992). “Rate of loudness growth for pure tones in normal and impaired hearing,” *J. Acoust. Soc. Am.* **93**, 966–975.
- Hellman, R. P., and Zwislowski, J. J. (1963). “Monaural loudness function at 1000 cps and interaural summation,” *J. Acoust. Soc. Am.* **35**, 856–865.
- Levitt, H. (1971). “Transformed up-down methods in psychoacoustics,” *J. Acoust. Soc. Am.* **49**, 467–477.
- Marozeau, J., and Florentine, M. (2007). “Loudness growth in individual listeners with hearing losses: A review,” *J. Acoust. Soc. Am.* **122**, EL81–EL87.
- Moore, B. C. J. (1998). *Cochlear Hearing Loss* (Whurr, London), pp. 1–332.
- Pratt, H., and Sohmer, H. (1977). “Correlations between psychophysical magnitude estimates and simultaneously obtained auditory nerve, brain stem and cortical responses to click stimuli in man,” *Electroencephalogr. Clin. Neurophysiol.* **43**, 802–812.
- Scharf, L. (1991). *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis* (Addison Wesley, Reading, U.K.), pp. 386.
- Serpanos, Y. C., O’Malley, H., and Gravel, J. S. (1997). “The relationship between loudness intensity functions and the click-ABR wave V latency,” *Ear Hear.* **18**, 409–419.
- Silva, I. (2009). “Estimation of postaverage SNR from evoked responses under nonstationary noise,” *IEEE Trans. Biomed. Eng.* **56**, 2123–2130.
- Silva, I., and Epstein, M. (2010). “Estimating loudness growth from tone-burst evoked responses,” *J. Acoust. Soc. Am.* **127**, 3629–3642.
- Stevens, J. C., and Guirao, M. (1964). “Individual loudness functions,” *J. Acoust. Soc. Am.* **36**, 2210–2213.
- Stevens, S. S. (1936). “A scale for the measurement of a psychological magnitude: Loudness,” *Psychol. Rev.* **43**, 405–416.
- Stevens, S. S. (1955). “The measurement of loudness,” *J. Acoust. Soc. Am.* **27**, 815–827.
- Stevens, S. S. (1957). “Concerning the form of the loudness function,” *J. Acoust. Soc. Am.* **29**, 603–606.
- Stevens, S. S. (1961). “To honor Fechner and repeal his law,” *Science* **13**, 80–86.
- Wilson, K. G., and Stelmack, R. M. (1982). “Power functions of loudness magnitude estimations and auditory brainstem evoked responses,” *Percept. Psychophys.* **31**, 561–565.
- Zwicker, E., and Fastl, H. (1999). *Psychoacoustics: Facts and Models* (Springer, New York), pp. 1–428.