

Categorical loudness scaling and equal-loudness contours in listeners with normal hearing and hearing loss

Daniel M. Rasetshwane,^{a)} Andrea C. Trevino, Jessa N. Gombert,^{b)}
Lauren Liebig-Trehearn,^{b)} Judy G. Kopun, Walt Jesteadt, Stephen T. Neely,
and Michael P. Gorga

Center for Hearing Research, Boys Town National Research Hospital, 555 North 30th Street, Omaha,
Nebraska 68131

(Received 16 October 2014; revised 10 March 2015; accepted 16 March 2015)

This study describes procedures for constructing equal-loudness contours (ELCs) in units of phons from categorical loudness scaling (CLS) data and characterizes the impact of hearing loss on these estimates of loudness. Additionally, this study developed a metric, level-dependent loudness loss, which uses CLS data to specify the deviation from normal loudness perception at various loudness levels and as function of frequency for an individual listener with hearing loss. CLS measurements were made in 87 participants with hearing loss and 61 participants with normal hearing. An assessment of the reliability of CLS measurements was conducted on a subset of the data. CLS measurements were reliable. There was a systematic increase in the slope of the low-level segment of the CLS functions with increase in the degree of hearing loss. ELCs derived from CLS measurements were similar to standardized ELCs (International Organization for Standardization, ISO 226:2003). The presence of hearing loss decreased the vertical spacing of the ELCs, reflecting loudness recruitment and reduced cochlear compression. Representing CLS data in phons may lead to wider acceptance of CLS measurements. Like the audiogram that specifies hearing loss at threshold, level-dependent loudness loss describes deficit for suprathreshold sounds. Such information may have implications for the fitting of hearing aids. © 2015 Acoustical Society of America.

[<http://dx.doi.org/10.1121/1.4916605>]

[ELP]

Pages: 1899–1913

I. INTRODUCTION

One factor that influences satisfaction with hearing aids for individuals with sensorineural hearing loss is the percept of loudness (e.g., Shi *et al.*, 2007; Blamey and Martin, 2009). Although current hearing-aid fitting strategies consider loudness perception in the prescription of gain, they typically do not use individual loudness measurements obtained across the dynamic range of hearing to set gain as a function of input level. Instead, models of loudness are used, in part, because of the difficulty in making reliable loudness measurements and the time and training such measurements require. Categorical loudness scaling (CLS) potentially circumvents these problems; however, CLS has been criticized because of continuing concerns related to the reliability of the measurements and because it uses arbitrary ordinal units [categorical units (CUs)] that are difficult to relate to standard units of loudness (sones) or loudness level (phons). The aims of this study were to (1) further assess the reliability of CLS measurements, (2) describe procedures for expressing CLS data that have been measured in CUs as loudness level in units of phons, (3) construct equal-loudness contours (ELCs), (4) characterize the impact of hearing loss on ELCs

derived from CLS measurements, and (5) present an approach that uses CLS data to specify the deviation in dB from normal loudness perception at various loudness levels for an individual with hearing loss. Representing CLS data in phons may lead to a wider acceptance of CLS measurements. The deficit from normal loudness perception for suprathreshold sounds, which we refer to as *level-dependent loudness loss*, may have implications for the fitting of hearing aids. This study extends previous studies by including a greater number of participants and a larger set of stimulus frequencies.

Loudness is defined as a listener's subjective response to the intensity of a sound (Scharf, 1978). The sensation of loudness primarily depends on the physical intensity of a sound [i.e., sound pressure level (SPL)]. However, loudness has been shown to also depend on the spectral (e.g., Garnier *et al.*, 2000; Anweiler and Verhey, 2006; Leibold *et al.*, 2007) and temporal (e.g., Zwislocki, 1969; Florentine *et al.*, 1996; Garnier *et al.*, 1999a) properties of a sound, as well as other factors. Several psychoacoustic procedures have been developed to measure loudness. Traditional methods include magnitude estimation, magnitude production, and cross-modality matching (e.g., Marks and Florentine, 2011). In magnitude estimation, the listener's task is to rate perceived loudness using a continuous and unbounded scale. In magnitude production, the listener's task is to adjust the intensity of a sound to achieve a loudness perception that is proportional to a specific number. Cross-modality matching is a process by which a listener adjusts the magnitude of a

^{a)}Author to whom correspondence should be addressed. Electronic mail: daniel.rasetshwane@boystown.org

^{b)}Current address: Department of Special Education and Communication Disorders, University of Nebraska-Lincoln, 301 Barkley Memorial Center, Lincoln, NE 68583.

physical quantity, such as string length or the brightness of a light, to match the loudness of a sound. When well designed, these methods can provide data that satisfy desirable properties of loudness measurement, such as internal consistency and transitivity of loudness scales (e.g., Marks and Florentine, 2011). However, these measurements are not appropriate for clinical application as they are time consuming, require extensive training, and may not be reproducible in individual listeners across test sessions. CLS may circumvent some of these problems, in that CLS (1) is easy to perform and relates to a listener's experience and informal descriptions of his or her loudness percepts, (2) requires minimal training, (3) is reliable, and (4) requires an amount of time that might be acceptable clinically (Allen *et al.*, 1990; Brand and Hohmann, 2002; Al-Salim *et al.*, 2010; Oetting *et al.*, 2014). These attributes may make CLS more attractive for clinical applications, compared to traditional methods for measuring loudness. In CLS, signals of different intensities (bounded between just above threshold and uncomfortably loud) are presented to a listener who is asked to rate the loudness of each signal using a series of loudness categories [e.g., Allen *et al.*, 1990; Brand and Hohmann, 2002; International Organization for Standardization (ISO), 2006; Al-Salim *et al.*, 2010]. The loudness categories are usually assigned meaningful descriptors, such as "very soft," "medium," and "very loud."

Despite its advantages, CLS has not been widely accepted for clinical applications, in part, because of concerns related to the reliability of the measurements (e.g., Elberling, 1999), its use of arbitrary units (CUs) that do not relate to standard units of loudness or loudness level, its use of bounded categories and its underestimation of the slope of loudness growth (Hellman, 1999). Another criticism is that CLS lacks the properties of a ratio scale (Stevens, 1946), i.e., it does not quantify how much louder one sound is compared to another; it only provides information regarding the ranking of sounds from softest to loudest. While previous studies have demonstrated the reliability of CLS measurements (e.g., Robinson and Gatehouse, 1996; Cox *et al.*, 1997; Rasmussen *et al.*, 1998; Al-Salim *et al.*, 2010), these studies focused on group data, instead of data from individual participants. And last, although several researchers have considered using CLS for fitting hearing aids, it remains unclear how these data should be incorporated into the fitting and, in cases where it is used, only limited benefit has been demonstrated (e.g., Cox, 1995; Ricketts, 1996). Consequently, further evaluation of CLS measurements is required if they are going to be accepted for clinical use.

To address the above concerns, this study further assessed the reliability of CLS, i.e., the repeatability of measurements across test sessions, by comparing the similarity in the SPL required to elicit the same loudness judgment (i.e., CU) across two independent test sessions at the same test frequency. This assessment was done on an individual basis at more frequencies than has been done in previous studies (e.g., Robinson and Gatehouse, 1996; Cox *et al.*, 1997; Rasmussen *et al.*, 1998; Al-Salim *et al.*, 2010). Repeatability of CLS measurements on an individual basis is

important if the measurements are to be adopted for clinical use.

This study also describes a procedure for converting CLS data from loudness in CUs to loudness level in phons. The phon scale is the ISO standard unit of loudness level for pure tones and is defined in terms of the SPL of a frontally incident 1-kHz tone (ISO, 2003). A tone of any frequency is said to have a loudness level of N phons if it is perceived by a listener with normal hearing to be as loud as a 1-kHz tone at N dB SPL (where N phons and N dB SPL are equal). We use this relationship to transform loudness in CUs to loudness level in phons by viewing CLS as a loudness-matching procedure. Having CLS data, converted to phons, for a wide range of frequencies allows for the construction of ELCs. ELCs are functions in the SPL/frequency domain that connect points whose coordinates represent pure tones judged to be equally loud (ISO, 2003). Our technique for conversion of CLS data to phons and construction of ELCs for participants with normal hearing is similar to the procedure used by Heeren *et al.* (2013). Constructing ELCs allows for a comparison of CLS data to loudness level measured using traditional loudness-matching methods (e.g., ISO, 2003; Suzuki and Takeshima, 2004).

The effects of hearing loss on CLS measurements have been described previously (e.g., Ricketts and Bentler, 1996; Brand and Hohmann, 2001; Al-Salim *et al.*, 2010). In the present study, we expand this characterization in two ways. First, we measured CLS functions in a large group of individuals with normal hearing and with mild-to-severe hearing loss across a wide range of frequencies. Just as CLS data are used to construct ELCs for participants with normal hearing, CLS functions for listeners with hearing loss are used to construct ELCs, which allows for a description of the ways in which ELCs are affected by hearing loss. Second, we describe an approach for characterizing the deviation from normal loudness for individual listeners with hearing loss. This level-dependent loudness loss describes the supra-threshold consequences of hearing loss in a level-dependent manner and may inform loudness-based fittings of hearing aids.

Listeners with sensorineural hearing loss typically present with loudness recruitment, a phenomenon in which the rate of loudness growth with sound intensity is more rapid than normal (e.g., Steinberg and Gardner, 1937; Stevens and Guirao, 1967; Scharf, 1978; Hellman and Meiselman, 1993). Because wide-dynamic-range compression applies less gain as the input level increases, it can be used to compensate for abnormal growth of loudness. Although wide-dynamic-range compression may counteract loudness recruitment, the percept of inappropriate loudness still remains one of the factors that causes dissatisfaction with hearing aids (e.g., Shi *et al.*, 2007; Blamey and Martin, 2009), perhaps because mean data, as opposed to individual data, are used when setting the characteristics of wide-dynamic-range compression for individual patients. Most current hearing-aid fitting strategies incorporate consideration of loudness perception in some form. The Desired Sensation Level algorithm aims to produce near-normal loudness growth (Cornelisse *et al.*, 1995). The National Acoustic Laboratories algorithm aims to

maximize the intelligibility of speech without exceeding normal loudness (Byrne *et al.*, 2001). Cambridge Loudness Equalization aims to amplify speech to provide equal loudness per critical band (Moore *et al.*, 1999). However, all of these fitting strategies use hearing threshold levels, group average data, and mathematical models of loudness or speech intelligibility to determine gain and output levels. They do not use individual measurements of loudness, in spite of the fact that rate of loudness growth is not always predictable from the degree of hearing loss (Ricketts and Bentler, 1996; Hellman, 1999). Individual measurements of loudness are not typically used, in part, because of the difficulty in making measurements, as discussed earlier. When measures of loudness are made, they are used to resolve estimates of loudness discomfort level, so that the presentation of uncomfortably loud sounds can be avoided (e.g., Cornelisse *et al.*, 1995). Most sounds that people listen to, however, are below typical loudness discomfort levels. Characterization of the deviation from normal loudness perception for levels between hearing threshold levels and loudness discomfort level may lead to hearing-aid prescription strategies that ameliorate issues related to loudness perception for the entire dynamic range of hearing. As a first step toward this goal, this study describes an approach that uses individual loudness data to specify the deviation from normal loudness perception as a function of frequency and loudness level. Level-dependent loudness loss specifies the deviation in dB from normal loudness perception for suprathreshold sounds for an individual listener with hearing loss. We will describe the application of level-dependent loudness loss for generating level-dependent gain for a hearing aid. A concept that is similar to level-dependent loudness loss was introduced earlier by Kollmeier *et al.* (1993). The current study extends this earlier study and other studies that have proposed hearing-aid fitting strategies using loudness data (e.g., Cox, 1995; Ricketts, 1996) by specifying level-dependent loudness loss for broad regions of the dynamic range of hearing, and for a wide range of frequencies in a large number of participants.

II. METHODS

A. Participants

A total of 148 participants took part in this study. There were 87 (50 female) participants with hearing loss (mean age = 55.0, SD = 17.6, range = 13–75), and 61 (45 female) participants with normal hearing (mean age = 28.9, SD = 10.7, range = 11–53). All participants were assessed using standard audiometry. Pure-tone air-conduction thresholds were measured at octave (0.25, 0.5, 1, 2, 4, and 8 kHz) and inter-octave frequencies (0.75, 1.5, 3, and 6 kHz), while pure-tone bone-conduction thresholds were measured at octave frequencies from 0.25 to 4 kHz. Thresholds were measured in 5-dB steps, following standard clinical procedures. Participants with air-conduction thresholds ≤ 15 dB hearing level (HL) at all frequencies were considered to have normal hearing, although the group with normal hearing was divided further into those with thresholds less than 5 dB HL and those with thresholds between 5 and 15 dB HL

(see below). Participants with thresholds >15 dB HL at one or more test frequencies were considered to have hearing loss. Participants were excluded from the study if air-bone gaps were >10 dB at any frequency. Middle-ear function was assessed using 226-Hz tympanometry, and participants were included if static compliance was between 0.3 and 2.5 cm³ and middle-ear pressure ranged from +50 to -100 daPa. Participants not meeting these criteria were excluded from the study. This study was conducted under an approved Institutional Review Board protocol and informed consent was obtained from all participants.

A subset of the participants (5 with normal hearing and 17 with hearing loss) returned for a second visit and their data were used to assess test-retest repeatability of CLS measurements. The range of time between visits was 2–79 days, with mean = 25.6 and SD = 18.2 days.

B. Equipment

Stimuli for CLS measurements were generated using a 24-bit sound card (Track16, MOTU, Cambridge, MA). The stimuli were routed via a headphone amplifier (HP4, PreSonus, Baton Rouge, LA) and were presented unilaterally to the participant's test ear with a headphone (HD-25-1 II, Sennheiser, Ireland). The headphone amplifier allowed for production of stimulus levels up to 110 dB SPL (the maximum presentation level). The Sennheiser HD-25-1 II headphones do not deliver a flat response up to 8 kHz, but instead provide enhancement for frequencies above 6 kHz. The frequency response of the headphones was measured using a sound level meter (System 824; Larson Davis, Provo, UT) and used to derive SPL corrections that were subsequently applied to the CLS measurements. Custom-designed software (Behavioral Auditory Research Tests, version 2.3.32) was used to control stimulus delivery and record the responses of the participants. In addition to the headphone correction, a correction for the acoustics of the pinna and ear canal was measured using Knowles Electronics Manikin for Acoustic Research (KEMAR, Knowles Electronics, Itasca, IL) and a sound level meter. Although this correction is not a true free-field correction (e.g., Shaw, 1965), it accounts for most of the frequency response-shaping propagation path of the stimulus, missing only the propagation from external loudspeaker to pinna. This correction facilitates comparison of our loudness measurements to the ISO (2003) standard, and other loudness measurements which were made with a free-field sound source.

C. Measurements and stimuli

The CLS procedure determined the level of pure tones that corresponded to different loudness categories using an adaptive procedure (Brand and Hohmann, 2002; ISO, 2006; Al-Salim *et al.*, 2010). A total of 11 loudness categories were used, with 7 of these categories assigned meaningful labels (“can’t hear,” “very soft,” “soft,” “medium,” “loud,” “very loud,” and “too loud”). The categories were displayed using colored horizontal bars with increasing length from the softest level to the loudest level, as shown in Fig. 1. The labels used at the boundary categories, “can’t hear” and “too loud,” are different from those used in the ISO (2006)

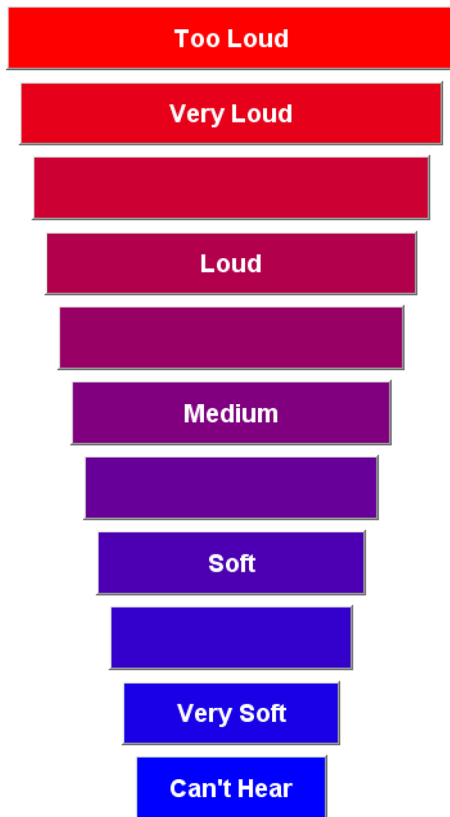


FIG. 1. (Color online) Categorical loudness scale with 11 response categories displayed on a computer monitor used by participants to rate the loudness of the signal. The horizontal bars increase in width from the softest level to the loudest level.

and the Heeren *et al.* (2013) study. The ISO recommends “not heard” and “extremely loud,” while Heeren *et al.* used “inaudible” and “too loud.” The implications of these differences will be discussed when data are presented. However, it is worthwhile to note that the ISO (2006) is open to the use of different labels for the categories, including symbols. Participants used a computer mouse to click on the category that best matched their loudness perception. Participants were encouraged to use both labeled and unlabeled bars. The CLS procedure included two stages. In the first stage, the dynamic range of the participant was determined by presenting two sequences of stimuli, one sequence ascending in level and the other descending in level. The lower end of a participant’s dynamic range was based on the last audible level of the descending sequence, while the upper end was based on the last level of the ascending sequence that was not judged as “too loud.” In the second stage, stimuli were presented at 11 levels equally spaced within the participant’s dynamic range, and the participant was asked to rate the loudness of the stimuli using loudness categories, as described above. Additional stimulus presentations were made if any two consecutive levels were judged to be >10 CUs apart, in order to better define the shape of the function. The second stage of the CLS procedure was repeated three times using the same levels. Thus, there were a total of at least 33 presentations in the second stage alone. The order of presentation of the stimulus levels within each of the three

trials was randomized. The starting presentation level was fixed at 60 dB SPL for participants with normal hearing. For participants with hearing loss, the starting level was half way between their threshold and the maximum presentation level. For the purpose of numerical representation, the 11 loudness categories were assigned CUs from 0 (can’t hear) to 50 (too loud) in steps of 5. The pure-tone stimuli were 1000 ms in duration with rise/fall times of 20 ms. Data collection at a single frequency took approximately 3–4 min to complete.

CLS measurements were made monaurally and separately for each frequency. Following a practice run at 1.25 kHz, measurements were made at a total of ten octave and inter-octave frequencies in participants with normal hearing. For all participants with hearing loss, measurements were made at octave frequencies. Additional measurements at inter-octave frequencies were made in participants with hearing loss if their behavioral thresholds at two consecutive octave frequencies differed by 20 dB or more. As a result, some participants with hearing loss contributed data at six frequencies, while others contributed data at as many as ten frequencies. Fewer measurements were made in participants with hearing loss in order to reduce the amount of time required for data collection. For those who participated in repeated measurements, test and retest data were collected on a separate day for the same frequencies that were tested initially.

If both ears met the inclusion criteria for participants with normal hearing, the ear with better hearing was selected for testing. If both ears met the inclusion criteria for participants with hearing loss, the ear with audiometric thresholds in the mild-to-moderate range was tested. For both groups, if the two ears had similar audiometric thresholds, the test ear was selected randomly. Data were collected for only one ear per participant. Overall, data were collected in 49 right and 38 left ears with hearing loss and 33 right and 28 left ears with normal hearing.

D. Data analysis

1. CLS function

For each participant and at each frequency, CLS data from the three presentations were analyzed to obtain a CLS function (i.e., loudness in CUs as a function of dB SPL). The CLS-function analysis followed the procedure described by Al-Salim *et al.* (2010) and Rasetshwane *et al.* (2013). In this analysis procedure, outliers were removed and the median SPL for each CU was calculated.¹ Two definitions of outliers, applied at different stages of the analysis, were used. In the first stage, data were collapsed across the three presentations and then, for each CU, data that deviated by more than 12 dB from the median SPL for the three presentations were considered outliers and removed from further analysis. The remaining data points were used to obtain a median level for a given CU. Note that, since the median SPL was calculated for each CU, some participants assigned multiple levels to a particular CU and hardly used some other CUs.

The second definition of outliers was intended to make the CLS function monotonic because the perceptual correlate of an increase in stimulus intensity is expected to be an

increase in loudness. Thus, the change in level between two successive CUs, when going from low to high CUs, was required to be a minimum of +1 dB. Any data points that did not meet this requirement were considered outliers and excluded. The first definition of outliers removed 16% of the data and the second definition removed an additional 10% of the data after the first outlier rule was applied. Finally, a CLS function was obtained from the median CLS data by fitting a model loudness function to the data. The model loudness function consisted of two linear functions with independent slopes, one for the portion of the data from 5 to 25 CUs and the other for the portion from 25 to 45 CUs. Justification for this two-segment approach is provided by Al-Salim *et al.* (2010) and Brand and Hohmann (2002), in which they observed a break in the slope of the CLS function at 25 CUs. The two-part function was fit to the data using a least-square fit approach, similar to that of Brand and Hohmann (2002). Data for CU = 0 (not heard) and CU = 50 (too loud) were not used in the fit because the SPLs for these categories are unconstrained. That is, a decrease in the stimulus level below threshold by any amount will still result in a rating of “can’t hear.” Equivalently, an increase in the stimulus level by any amount above a level that was rated “too loud” would still result in a rating of “too loud.” Note that because of the fitting procedure, the CLS function can include dB SPL values that are greater than the maximum output of our sound delivery system (110 dB SPL). Following previous studies (e.g., Anweiler and Verhey, 2006; Al-Salim *et al.*, 2010; Rasetshwane *et al.*, 2013; Heeren *et al.*, 2013; Oetting *et al.*, 2014), all subsequent analyses were performed on the fitted CLS function.

Data from all participants (normal hearing and hearing loss) were placed into seven categories based on audiometric-threshold levels (dB HL) at each frequency. Grouping the data by audiometric threshold allows us to characterize the effect of hearing loss on CLS and ELCs derived from CLS data. The ranges for the seven groups were -10 to 0, 5 to 15, 20 to 30, 35 to 45, 50 to 60, 65 to 75, and 80 to 90 dB HL. The number of participants contributing data to each threshold category and test frequency is shown in Table I. The column sums for octave frequencies are not all equal to 148 (the total number of participants) because a few participants were unable to complete the study. As expected, there were more participants contributing data to the first two threshold categories, -10 to

0 and 5 to 15 dB HL, in the low frequencies, as some of the participants with hearing loss had thresholds in the normal range at these frequencies. There were also few participants in the last threshold category, 80–90 dB HL, regardless of frequency.

2. Repeatability

CLS test-retest reliability was assessed by comparing the SPL required to elicit the same loudness perception across two test sessions. This assessment was performed on individual loudness data for the participants who contributed data on two separate visits (5 normal hearing and 17 hearing loss). Reliability was assessed using four metrics; (1) mean (across CUs) of the signed differences (*Mean-SgnDiff*), (2) standard deviation (across CUs) of the signed differences (*SD-SgnDiff*), (3) Cronbach’s α (Cronbach, 1951), and (4) standard error of measurement (*Sm*). Cronbach’s α is a coefficient of internal consistency that can be used as an estimate of reliability. Cronbach’s α is calculated as

$$\alpha = \frac{K}{K-1} \left(1 - \frac{\sum_{i=1}^K \sigma_{Y_i}^2}{\sigma_X^2} \right), \quad (1)$$

where σ_X^2 is the variance of the total measurements, $\sigma_{Y_i}^2$ is the variance of measurement i , and K is the number of measurements. α ranges from 0 to 1, where a value of $\alpha = 0$ indicates that the measurements are totally unrelated and a value of $\alpha = 1$ indicates perfect reliability. A value of $\alpha \geq 0.85$ is considered good reliability. *Sm* is defined as the standard deviation of the measurement error and is calculated as

$$Sm = SD\sqrt{1 - \alpha}, \quad (2)$$

where α is Cronbach’s α , and *SD* is the standard deviation of the SPLs across participants for a given test condition (e.g., Wagner *et al.*, 2008). A *Sm* > 2 dB is considered indicative of significant deviation in the measurements across repetitions because $\pm 1.96 Sm$ is equivalent to the 95% confidence interval. All four metrics were calculated separately for each frequency and participant, with the mean and standard deviation of the differences and the *SD* in Eq. (2) calculated for each participant. The calculation of the metrics was based on comparison of fitted CLS functions from the test and retest sessions. Because data were not collected at inter-octave frequencies in all participants with hearing loss, reliability was assessed only at octave frequencies, i.e., 0.25, 0.5, 1, 2, 4, and 8 kHz.

3. CLS to ELC conversion

The process of converting CLS data in CUs to loudness level in phons involves several steps. In the first step, the relationship between audiometric threshold and stimulus level in dB SPL for each loudness category (CU) was determined by considering the data from all participants together. A function was calculated separately for each loudness category by fitting the data with a simple linear regression,

TABLE I. Number of participants contributing data at each hearing-loss group and test frequency.

Thresholds (dB HL)	Frequency (kHz)										Total
	0.25	0.5	0.75	1	1.5	2	3	4	6	8	
-10, -5, 0	23	29	20	31	29	34	33	27	12	10	248
5, 10, 15	79	66	53	58	44	47	37	44	53	54	535
20, 25, 30	29	21	11	15	19	18	13	9	8	13	156
35, 40, 45	10	22	3	22	8	24	12	20	12	11	144
50, 55, 60	5	5	4	14	1	17	7	27	9	20	109
65, 70, 75	2	3	3	5	2	5	6	17	5	29	77
80, 85, 90	0	1	1	1	2	3	0	1	2	7	18
Total	148	147	95	146	105	148	108	145	101	144	

resulting in a set of nine representative functions (CU values of 5–45) for each of ten frequencies. For a given frequency, the functions described the relationship between stimulus level in dB SPL and audiometric threshold in dB HL for each of the nine CU values. Each linear fit represents an equal-loudness function with audiometric threshold as the independent variable. Recall that the phon is defined as the loudness level of a 1-kHz tone at some SPL, and that tones of different frequencies equal in loudness to the 1-kHz tone are defined as having the same loudness level (in phons). Thus, the intercepts of these linear fits with 0 dB HL at 1 kHz describe the relationship between stimulus level in dB SPL and loudness in CU for listeners with an audiometric threshold of 0 dB HL; these intercepts define the CU-to-phon conversion.

Next, by treating CLS as a loudness-matching procedure (i.e., by assuming, for example, that 15 CU at 1 kHz is as loud as 15 CU at 6 kHz), the CU-to-phon conversion function derived for 0 dB HL was applied to convert CLS data at other frequencies from loudness in CUs to loudness level in phons. Our treatment of CLS as loudness-matching procedure follows several previous studies (e.g., Anweiler and Verhey, 2006; Al-Salim *et al.*, 2010; Heeren *et al.*, 2013). A plot of the levels that correspond to equal-loudness perception as a function of frequency results in ELCs for a group with audiometric thresholds of 0 dB HL. Finally, by evaluating the intercepts of the linear functions relating stimulus level to audiometric threshold with dB HL values greater than zero, and by applying the CU-to-phon conversion function obtained at 0 dB HL, ELCs for different degrees of hearing loss were obtained. We evaluated ELCs that represent 20, 40, and 60 dB HL flat losses. We elected to use linear regression for the CU-to-phon conversion because it allowed us to utilize loudness data from all participants and to exploit the relationship between the SPL associated with a given CLS category and audiometric threshold. If the median or mean associated with a given CLS category was used, the conversion function would be limited to data from only participants with normal hearing. The procedures for converting CLS data to loudness level in phons and for constructing ELCs will be described in greater detail in Sec. III when CLS data are presented.

4. Level-dependent loudness loss

As described earlier, we define level-dependent loudness loss for listeners with hearing loss as their difference in dB from normal loudness perception. ELCs for 0 dB HL,

obtained as described above, were used as the reference for normal loudness perception. The first step in calculating the level-dependent loudness loss for a participant with hearing loss is to construct ELCs for the nine loudness categories corresponding to CUs of 5–45 from their CLS data. Level-dependent loudness loss is then calculated as the dB difference between the ELCs for the individual with hearing loss and the normal-hearing reference for matched CUs. Audiometric thresholds for the participant with hearing loss were also included as part of the description of level-dependent loudness loss because these thresholds can be interpreted as representing the deficit from normal loudness perception for just audible sounds. For this purpose, 0 dB HL is considered as normal hearing. Our characterization of level-dependent loudness loss may be interpreted as the gain that needs to be applied as a function of frequency and level (or loudness category) to achieve normal loudness perception for tones.

III. RESULTS

A. Repeatability

One aim of this study was to evaluate the test-retest reliability of CLS measurements on an individual basis. To achieve this aim, four metrics—(1) mean of the signed differences (*Mean-SgnDiff*), (2) standard deviation of the signed differences (*SD-SgnDiff*), (3) Cronbach's α , and (4) standard error of measurement (*Sm*)—were analyzed in terms of their cumulative distributions. Recall that the calculations of *Mean-SgnDiff*, *SD-SgnDiff* and Cronbach's α were based on data from the test and retest data for each individual participant. The SD used for the calculation of *Sm* [see Eq. (2)] was calculated across participants, separately for each frequency and loudness category (CU). The resulting *Sm* was then averaged across CUs. The metrics are shown in Fig. 2 in separate panels, as indicated by the abscissa labels. For each metric, separate cumulative distributions are shown for each octave frequency (solid lines), together with a cumulative distribution that is based on data from all frequencies (dashed line). Across all frequencies, the *Mean-SgnDiff* had an interquartile range (IQR) of -3.57 to 2.88 dB (mean = -0.53 , SD = 5.00 dB), *SD-SgnDiff* had an IQR of 2.04 to 5.36 dB (mean = 4.22 , SD = 2.88 dB), and Cronbach's α had an IQR of 0.98 to 1 (mean = 0.98 , SD = 0.02). In fact, the smallest value of Cronbach's α was 0.86 (at 4 kHz) and is greater than the critical value of 0.85 , which is considered the criterion for reliable data. Across all frequencies, *Sm* had

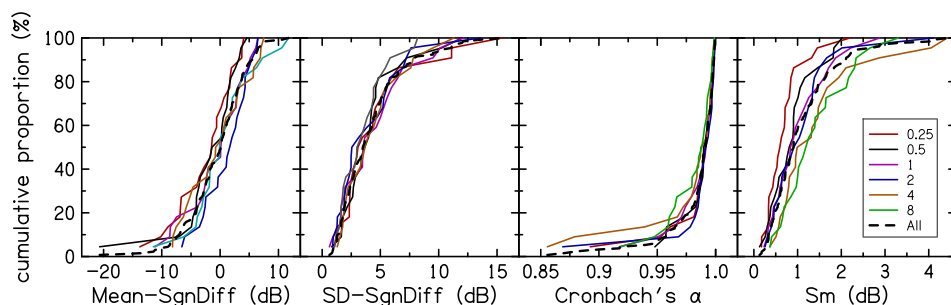


FIG. 2. (Color online) Reliability analyses using cumulative distributions of mean of signed differences (*Mean-SgnDiff*), SD of signed differences (*SD-SgnDiff*), Cronbach's α and standard error of measurement (*Sm*), as indicated in the abscissa labels. Separate cumulative distributions are shown for each test frequency (solid lines), together with a cumulative distribution that is based on data from all frequencies (dashed line).

an IQR of 0.57 to 1.42 dB, (mean = 1.10, SD = 0.79 dB). The cumulative distributions for the different frequencies were similar for each metric indicating that there was general consistency in reliability across frequency. For the analysis with data collapsed across frequency, Cronbach's α indicated that every participant produced reliable CLS data, as all values of α were greater than the critical value of 0.85. In contrast, the Sm suggested that 11.5% of the participants did not have reliable data, with $Sm > 2$, for these participants.

B. CLS and ELC

Figure 3 shows median CLS functions for the seven audiometric-threshold (hearing-loss) groups with data for a different frequency represented in each panel. The parameter in each panel is hearing-loss category (in dB HL), as indicated in the insets in the rightmost column. The CLS functions had shallow portions from near threshold to moderate levels and then steep portions at higher stimulus levels, which is typical of CLS functions (e.g., Ricketts and Bentler, 1996; Garnier *et al.*, 1999b; Brand and Hohmann 2001; Al-Salim *et al.*, 2010). As hearing-loss magnitude increased, the overall slopes of the CLS functions became steeper, demonstrating a reduction in dynamic range. This pattern was observed at all frequencies. Also, the shapes of these CLS functions for a particular hearing loss category were similar across frequency. The greatest differences among hearing-loss categories were observed for low-level inputs, and decreased as input level increased. That is, degree of hearing loss had a larger impact on loudness perception at low levels, compared to the impact at higher levels. This pattern is consistent with previously described CLS functions from participants with normal hearing and hearing loss (e.g., Ricketts and Bentler, 1996; Garnier *et al.*, 1999b; Al-Salim *et al.*, 2010) and previous demonstrations of loudness recruitment (e.g., Steinberg and Gardner, 1937; Stevens and Guirao, 1967; Scharf, 1978; Hellman and Meiselman, 1993). That is, loudness perceptions for high-level stimuli were less

dependent on threshold, compared to loudness perceptions for low-level stimuli. It is interesting to note that although the CLS functions for the -10 to 0 and 5 to 15 dB HL categories were similar; they were not identical, with the data from the 5–15 dB HL category shifted slightly to the right (i.e., higher stimulus levels). Thus, even within the range of normal hearing, differences were observed in loudness perception related to threshold.

To provide a sense of the amount of variability in CLS data, Fig. 4 shows the interquartile range (IQR, darker shading) and the 5th to 95th percentile range (lighter shading) of CLS functions at 1 kHz for six different hearing-loss categories. The median CLS functions are shown using black solid lines. Data for the 80–90 dB HL category are not shown because data from only one participant was available in this category, and thus variability cannot be calculated. Likewise, the apparently reduced variability for the 65–75 dB HL group should be viewed cautiously, as the number of participants contributing data to this condition was small ($N=5$; see Table I). The IQR of the CLS functions for the two categories that are within the normal range (i.e., -10 to 0 dB HL and 5 to 15 dB HL) varied from 13 to 21 dB and 10 to 18 dB, respectively. This amount of variability was slightly less than previously reported at 1 kHz for participants with normal hearing (e.g., Heeren *et al.*, 2013; ISO, 2006). Specifically, Heeren *et al.* (2013) reported CLS functions with an IQR that varied from approximately 20 to 26 dB, and the IQR for ISO 16832 varied from approximately 20 to 28 dB.² The IQR for our CLS functions were unchanged when the data were re-analyzed without removal of outliers, perhaps because of the large dataset.³ Therefore, removal of outliers did not have an effect on the CLS data when analyzed for the entire population. However, outliers affect individual participant data, as they can result in a non-monotonic CLS function.

Figure 5 plots the median stimulus level in dB SPL of the CLS functions as a function of audiometric threshold in dB HL with each curve representing a particular CU, as indicated by the insets in the rightmost panels. Data for different

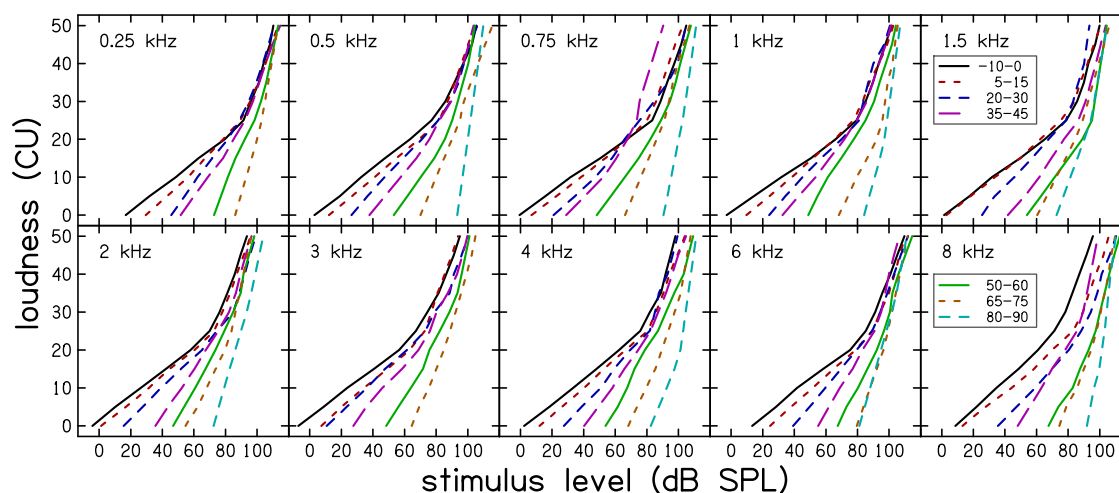


FIG. 3. (Color online) Median CLS functions for different hearing loss categories. Data for the ten test frequencies are shown in separate panels. The ranges of hearing loss for each group are shown in the right panels.

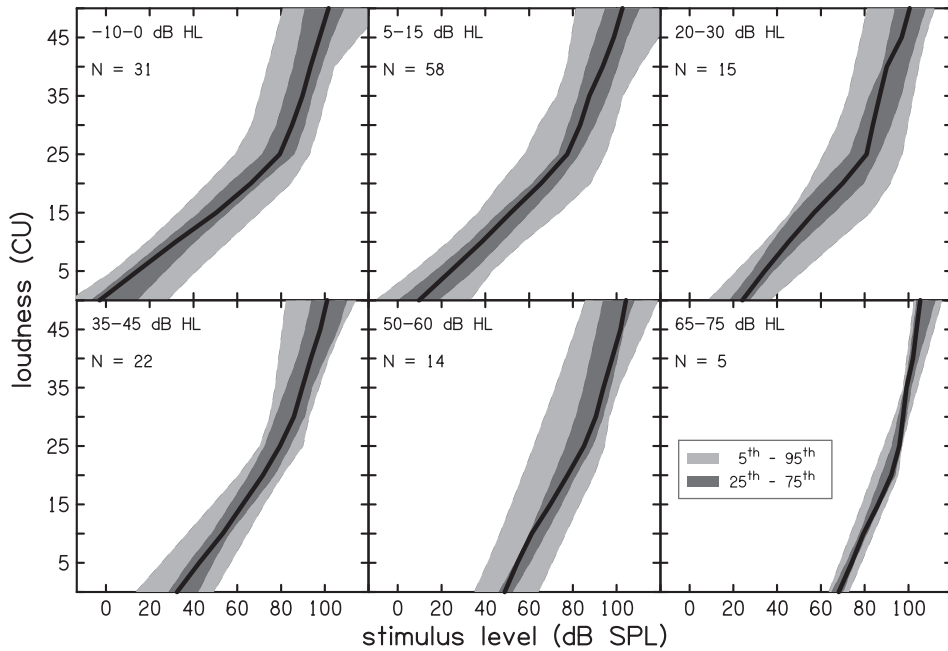


FIG. 4. Median CLS function (solid line) at 1 kHz. Data for the different hearing loss categories are shown in different panels. Lighter shading indicates 5th to 95th percentile range and darker shading indicates 25th to 75th percentile (interquartile) range, as shown in the inset in the bottom right panel. The number of participants contributing data at each hearing loss category is shown in the panels.

frequencies are shown in separate panels. While only medians of the CLS functions are plotted as symbols (for ease of visualization), the lines within each panel represent linear fits to the individual CLS functions for each of nine loudness categories (i.e., CU of 5–45). Because each linear fit within a panel represents the level necessary for a specific loudness category, each line can be viewed as an equal-loudness function. The slopes of these functions are steepest for soft sounds (i.e., at low CUs), decreasing in slope as loudness increases (i.e., at high CUs). The effect of audiometric threshold on stimulus level for equal loudness decreased as loudness increased (e.g., the slope of the line for CU = 45, top line in each panel, is close to zero). This means (not unexpectedly) that high input levels result in

similar loudness judgments regardless of audiometric threshold. Similar patterns were observed at all frequencies. The analysis of Fig. 5 extends the analysis of Al-Salim (2010; see their Fig. 6), by including more frequencies and a larger sample of participants.

The linear functions are calculated as part of the first step in the conversion of CLS data in CUs to phons, as described Sec. II. The intercepts of the regression lines with 0 dB HL at 1 kHz were used to derive a CU-to-phon conversion function. This conversion function is shown in Fig. 6, where circles show the median loudness level in phons that corresponds to CUs of 5 to 45. Recall that phons are defined as the intensity in dB SPL of an equally loud 1 kHz tone. Thus, the ordinate of Fig. 6 could be labeled

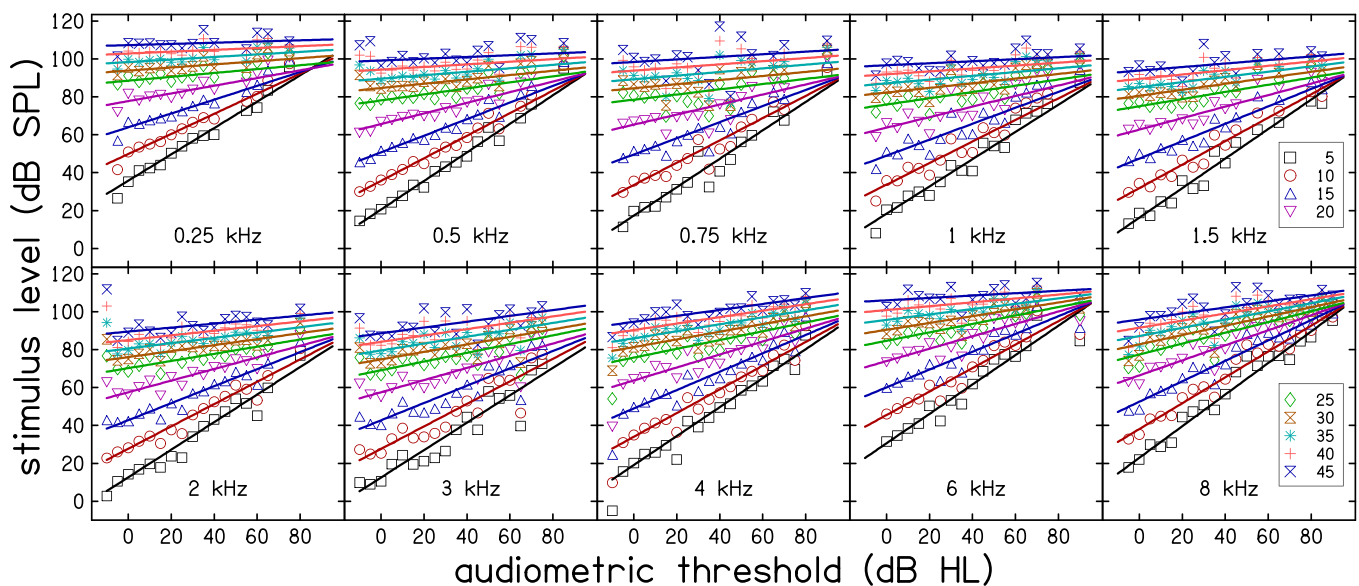


FIG. 5. (Color online) Median stimulus level in dB SPL as a function of audiometric threshold in dB HL for the loudness data. The parameter within each panel is CU, as indicated in the insets in the rightmost panels. Data for different frequencies are shown in separate panels. The lines within each panel are linear regression lines that were fit to the individual data for each of nine loudness categories (i.e., CU of 5–45). Each line is an equal-loudness function.

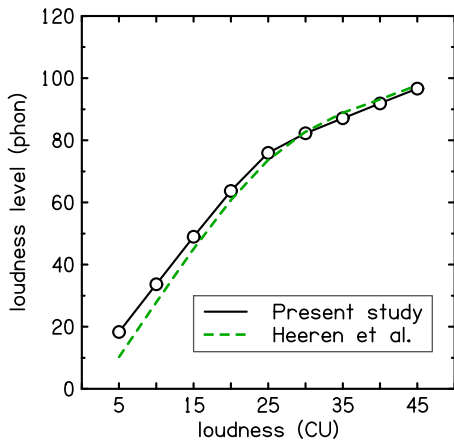


FIG. 6. (Color online) CU-to-phon conversion function for the present study (solid line with circle symbols) and for the study of Heeren *et al.* (2013; dashed line).

dB SPL at 1 kHz. Figure 6 also compares our CU-to-phon conversion function to that of Heeren *et al.* (2013), derived from their median CLS function at 1 kHz (dashed line). The current conversion function is similar to that of Heeren *et al.* The two functions are similar despite the fact that the labels used at the boundary categories differed. This suggests that the difference in the labels did not have an effect on the data.

Using the CU-to-phon conversion shown in Fig. 6, SPL-to-phon conversion functions were developed at other frequencies and plotted in Fig. 7. Figure 7 plots loudness level in phons as a function of stimulus level in dB SPL, with frequency as the parameter. The dashed diagonal line has a slope of one and represents the function relating loudness level in phons to absolute level in dB SPL at 1 kHz for a listener with an audiometric threshold of 0 dB HL. The left-most panel shows the conversion function for the 0 dB HL (i.e., normal hearing) group. As expected, our conversion function at 1 kHz for this group matches the ideal function for a listener with an audiometric threshold of 0 dB HL. The slopes of these functions are similar, suggesting that loudness level grows at similar rates for a wide range of frequencies in participants with normal hearing. There is a 20-dB range in intercepts with the abscissa, with the function for 0.25 kHz having the largest shift (to the right) from the dashed diagonal line. The difference in intercepts reflects the

dependence of threshold on frequency for individuals with normal hearing.

Figure 7 also shows the SPL-to-phon conversion functions for the groups with hearing loss, i.e., audiometric thresholds of 20, 40, and 60 dB HL (again, with frequency as the parameter within each panel). These functions were derived from the regression lines shown in Fig. 5 by solving for the dB SPL at which these lines intercept 20, 40, and 60 dB HL, and then applying the CU-to-phon conversion shown in Fig. 6. As hearing loss increased, the SPL-to-phon conversion functions became steeper, which is consistent with loudness recruitment. There was a change in slope of the SPL-to-phon conversion functions for the 40 and 60 dB HL groups that occurred around 90 dB SPL for the 40 dB HL group and around 95 dB SPL for the 60 dB HL group. This change of slope was also present in the 20 dB HL group, although it is not easily visualized in the figure. The shallower slope at high levels indicates that loudness level grows at a faster rate at low SPLs than at high SPLs. Note also that the differences among hearing-loss groups occur for low-level inputs, with more similar results for high-level inputs, which is an inevitable outcome, given the results shown in Figs. 3 and 5. The 20-dB range in intercepts with the abscissa that was observed in the SPL-to-phon conversion functions for the group with normal hearing is also present in the functions for the group with hearing loss, again with the function for 0.25 kHz having the largest shift (to the right) from the dashed diagonal line.

The functions shown in Fig. 7 can be used to generate ELCs, which are plots of level (dB SPL) as a function of frequency with loudness level in phons as the parameter. This is accomplished by determining values of the level in dB SPL where these functions intercept a specific loudness level in phons. Plotting the level (dB SPL) as a function of frequency produces an ELC for that specific loudness level (phon).

ELCs were constructed for loudness levels of 20–100 phons in 10-phon steps, and are shown in Fig. 8 for thresholds of 0 [Fig. 8(A)], 20 [Fig. 8(B)], 40 [Fig. 8(C)], and 60 dB HL [Fig. 8(D)]. Every other ELC for the 0 dB HL group [Fig. 8(A)] is labeled. ELCs in the other panels are not labeled, but the bottom curve represents the 20-phon ELC and the top curve represents the 100-phon ELC in all panels. The ELCs were more compressed as threshold increased,

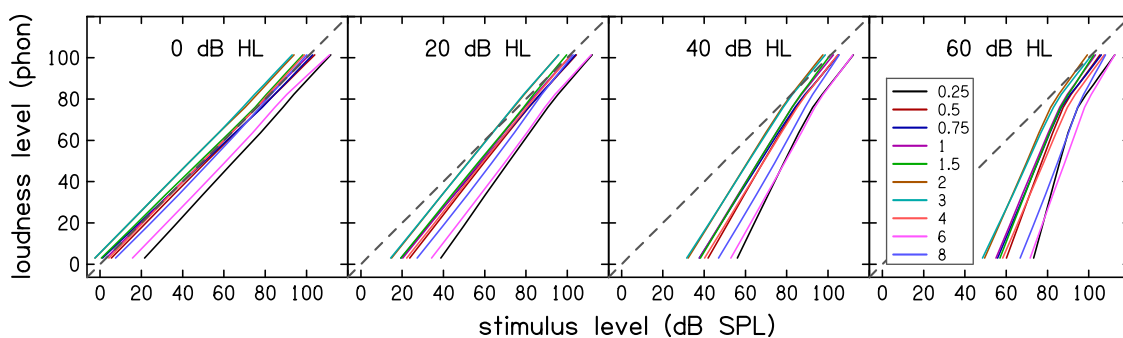


FIG. 7. (Color online) SPL-to-phon conversion functions (solid lines) for 0, 20, 40, and 60 dB HL. The dashed line in each panel represents the ideal conversion function for 0 dB HL at 1 kHz. Each panel shows conversion functions for the ten test frequencies using different colors, as indicated in the inset in the right panel.

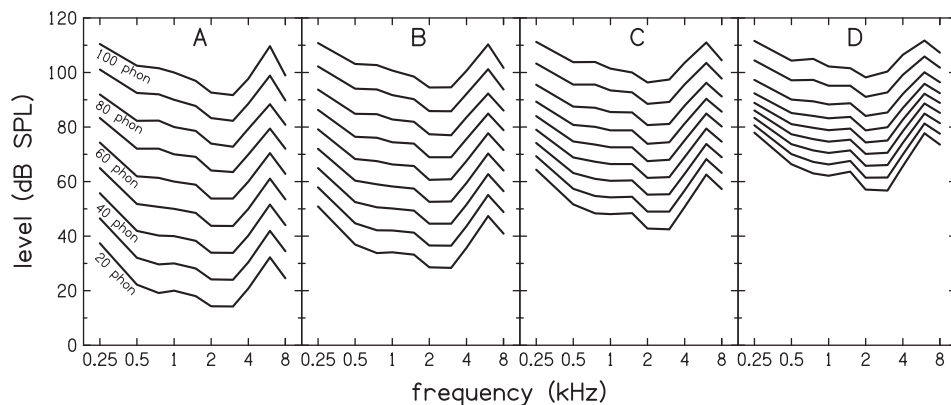


FIG. 8. ELCs for (A) 0, (B) 20, (C) 40, and (D) 60 dB HL constructed from CLS data. Each panel includes ELCs for 20 to 100 phons, in 10-phon steps.

indicating that loudness level grows more rapidly as audiometric threshold increases. The vertical spacing of the ELCs is relatively constant across frequency for threshold groups of 0, 20, and 40 dB HL, indicating that loudness level grows at approximately the same rate for all frequencies once threshold is taken into account. The ELCs for the 60 dB HL group are more closely spaced at low frequencies and at 8 kHz, suggesting that loudness grows more rapidly at these frequencies.

For comparison purposes, the ELCs for the 0 dB HL group [Fig. 8(A)] are reproduced in Fig. 9 for 20–100 phons, in 20-phon steps, together with ELCs from the International Standards for normal hearing (ISO, 2003), and ELCs from Heeren *et al.* (2013), which are also for listeners with normal hearing. Recall that the ELCs of Heeren *et al.* were measured using CLS. The present ELCs are in general agreement with those reported in the ISO standard and in Heeren *et al.* Differences are mainly observed in the extreme low and high frequencies and at high levels. At 0.25 kHz, the current contours for 20–80 phons are above the contours of the ISO

and Heeren *et al.* At 4 and 6 kHz, the current contours for 40–60 phons are also above the contours of the ISO and Heeren *et al.* In general, the current contours are more similar to those of Heeren *et al.* at mid-frequencies (1–3 kHz) than to the ISO contours. Another noteworthy difference between the CLS-measured ELCs and the ISO ELCs is that the former contours had a spacing that was more uniform across frequency, while the ISO contours had narrower spacing at low frequencies.

C. Level-dependent loudness loss

Figure 10 shows examples of level-dependent loudness loss for six participants with varying degrees and configurations of hearing loss. As described earlier, level-dependent loudness loss is calculated as the dB difference between the ELCs for the individual with hearing loss and the ELCs for normal hearing for matched CUs. For ease of visualization and to avoid clutter, loudness loss, compared to normal loudness perception, is shown only for three loudness categories (“very soft,” “medium,” and “very loud”). However, loudness loss can be obtained for any loudness category. Also shown in the figure are audiometric thresholds. The ordinate label in this case can be viewed as dB HL, as threshold and loudness loss are both based on comparisons to data from listeners with thresholds of 0 dB HL. The values plotted in Fig. 10 can also be viewed as the gain (in dB) that would be required to restore normal thresholds and normal loudness for these three loudness categories. These participants had greater loss at threshold and for soft sounds, indicating that more gain would be required to achieve a normal response for low-level inputs. In contrast, deviations from normal were reduced as loudness increased to the point that there was little or no loudness loss for “very loud” sounds (CU = 45). In fact, some participants (notably participants E and F) would not require gain at most frequencies for very loud sounds. The loudness-loss functions for these participants were occasionally equal to or less than zero, indicating that they perceived high-level stimuli as being louder than the perceived loudness for listeners with normal hearing for the equivalent input level. This observation may suggest that these participants are sensitive to loud sounds, although the extent to which these results were influenced by individual variability is not known.

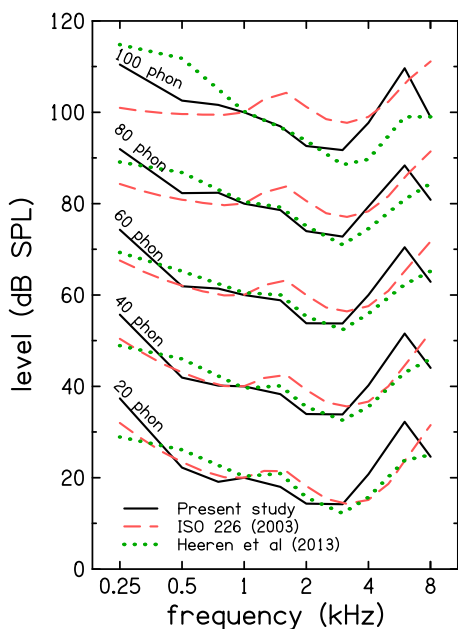


FIG. 9. (Color online) Comparison of the current ELCs for the 0 dB HL group (solid lines) to ELC of ISO (2003; dashed lines) and Heeren *et al.* (2013; dotted lines).

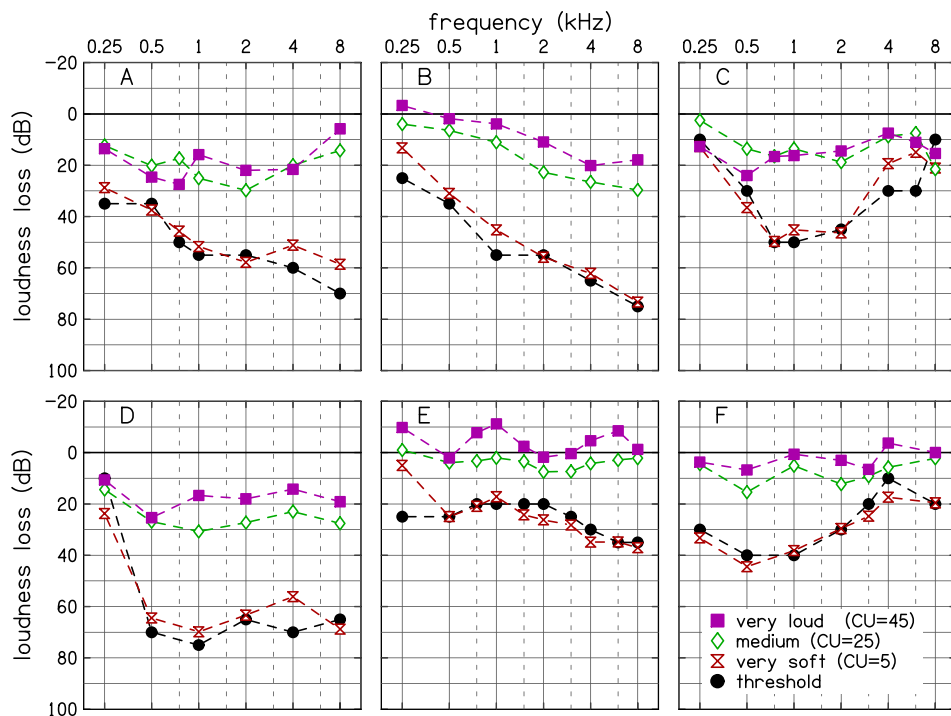


FIG. 10. (Color online) Level-dependent loudness loss for six participants with different degrees and configurations of hearing loss. In each panel, loudness loss or deviation from normal loudness perception are shown for “very soft,” “medium,” and “very loud” loudness categories, together with audiometric thresholds (see bottom right panel for figure legend).

IV. DISCUSSION

This study reassessed the test-retest reliability of CLS measurements, described CLS functions for both listeners with normal hearing and hearing loss, converted CLS data from loudness in CUs to loudness level in phons, constructed ELCs from the CLS data after conversion to phons, characterized the impact of hearing loss on derived ELCs, and presented an approach that uses CLS data to specify level-dependent loudness loss. Assessment of reliability was performed on individual data, instead of group data, as was the case in most previous studies. Repeatability of CLS measurements on an individual basis is important if the measurements are to be adopted for clinical use. For the analysis with data collapsed across frequency, Cronbach’s α indicated that every participant produced reliable CLS data, as all values of α were greater than the critical value of 0.85. In contrast, the Sm suggested that 11.5% of the participants did not have reliable data, with $Sm > 2$, for these participants. The reason for this discrepancy is that the Sm accounts for the variability across participants, as the SD across participants is included in its calculation [see Eq. (2)], while Cronbach’s α as used here does not account for the variability across participants. Overall, the present results suggest that CLS data are reliable, as differences in SPL between sessions for most participants were within acceptable limits (according to the four metrics used in this paper to assess reliability). There was consistency in reliability across frequency.

Several studies have evaluated the reliability of CLS measurements (Robinson and Gatehouse, 1996; Cox *et al.*, 1997; Rasmussen *et al.*, 1998; Al-Salim *et al.*, 2010), but focused on group data, instead of data from individual participants. For this reason, we had to convert our individual participant data to group data in order to compare the present results to those of the earlier studies. In the earlier studies,

reliability was typically assessed using mean and SD (both across loudness categories and participants) of signed differences. Figure 11 compares the *Mean-SgnDiff* (top panel) and *SD-SgnDiff* (bottom panel) of the present study (both calculated across CUs and participants) to results from Robinson and Gatehouse (1996), Rasmussen *et al.* (1998) and Al-Salim *et al.* (2010). At frequencies for which data existed in multiple studies, the present magnitudes of *Mean-SgnDiff* were similar to those reported by others except at 0.5 kHz, where our value of *Mean-SgnDiff* was larger. The bottom

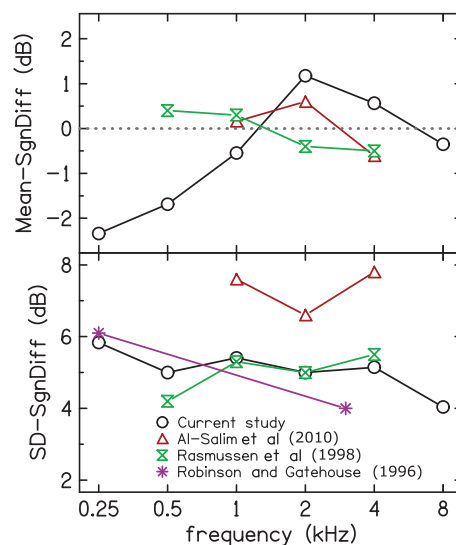


FIG. 11. (Color online) Comparison of reliability of the current CLS study (circle) to studies of Al-Salim *et al.* (2010; triangle), Rasmussen *et al.* (1998; hourglass) and Robinson and Gatehouse (1996; asterisk). The top panel shows the mean across participants of the mean signed differences (*Mean-SgnDiff*) and the bottom panel shows the mean of the standard deviation of the signed differences (*SD-SgnDiff*). For reference, *Mean-SgnDiff* = 0 dB is indicated in the top panel using a dotted line.

panel of Fig. 11 shows that the values of $SD\text{-}SgnDiff$ reported by Al-Salim *et al.* (2010) were larger than values from the other three studies, including our study, at frequencies that were common among studies. Overall, the CLS data reported in the present study demonstrate reliability that is similar to previous reports. However, our assessment of the reliability of group data is limited by our small sample size. Recall that test-retest data were only collected in 22 of the total 148 participants. It is of some interest that the frequencies for which the variability was greatest (0.25 and 0.5 kHz) are frequencies for which hearing loss is less common, compared to higher frequencies. For the higher frequencies for which hearing loss is more frequently encountered, less variability was observed. Efforts to further reduce individual variability may be necessary in order to optimize the clinical utility of the measurements.

It is possible that the time interval between test sessions may have influenced the repeatability of the CLS measurements. Since CLS is a behavioral procedure, one might assert that as the amount of time between test sessions increased, the responses of the participants became less reliable. To assess this possibility, the reliability of CLS measurements was assessed as a function of days between test sessions, which ranged from 2 to 79 days (mean = 25.6 days, SD = 18.2 days). For this analysis, the amount of variance explained by the days between test sessions was calculated for each metric and at each frequency using the R-statistic. The amount of variance explained by the days between test sessions ranged from 0.1% to 17.5% for $Mean\text{-}SgnDiff$, 0% to 1.7% for $SD\text{-}SgnDiff$, 0% to 5.6% for Cronbach's α , and 0% to 1.7% for Sm . These values represent up to a medium effect size for $Mean\text{-}SgnDiff$ and small effect sizes for the other three metrics (Cohen, 1992). However, none of the values of the R-statistic were statistically significant with $p > 0.05$ for all metrics and all frequencies. That is, there was no more than weak correlation between the reliability of the CLS measurements and amount of time between measurements. This is an encouraging observation, in that improvements in the measurement paradigm might be expected to reduce variability, and once variability is reduced, CLS results should be more stable over time.

CLS functions for individuals with normal hearing showed a shallow slope for low-to-moderate input levels and a steep slope for high input levels (Fig. 3). For individuals with hearing loss, the input level resulting in the softest loudness (smallest CUs) shifted to higher levels as hearing loss increased, reflecting threshold loss. Consequently, the slope of the CLS functions for low-to-moderate input levels became steeper as the degree of hearing loss increased. However, the slope of the function for high-level inputs was stable regardless of threshold. Stated another way, participants with normal hearing and those with hearing loss judged high-level inputs as equally loud. For the participants with thresholds of 65–75 dB HL and 80–90 dB HL, there was no obvious break point between low- and high-level segments in the CLS function, resulting in a slope that was uniform regardless of loudness. It is important to recall that most of the participants in this study had no worse than a mild-to-moderate hearing loss. With the exceptions of 4 and

8 kHz, there were few participants with hearing loss exceeding 60 dB HL at any frequency (see Table I). To a first approximation, it may be reasonable to assume that the hearing losses for the majority of participants were due to outer hair cell dysfunction (e.g., Liberman and Dodds, 1984). In humans, it is difficult to quantify the proportions of sensory cell loss (inner vs outer hair cells). However, in psychophysical studies in which attempts were made to quantify the proportion of the hearing loss that could be attributed to inner and outer hair cell dysfunction, (e.g., Lopez-Poveda and Johannesen, 2012; Johannesen *et al.*, 2014), the majority (60%–70%) of the loss was attributed to outer hair cell dysfunction. With greater degrees of hearing loss (and, presumably damage to inner hair cells), there might be less peripheral output, and perhaps decreases in loudness perception, compared to listener with normal hearing for high-level sounds. CLS measurements in participants with no worse than mild-to-moderate hearing loss perhaps provide a reference for future work, in which these same techniques are applied with participants having greater degrees of hearing loss.

The changes in the CLS functions with hearing loss observed in the present study were expected and similar to previous reports (e.g., Ricketts and Bentler, 1996; Garnier *et al.*, 1999b; Brand and Hohmann, 2001; Al-Salim *et al.*, 2010). In combination, these results demonstrate that CLS data are capable of showing loudness recruitment, like other measures of loudness (e.g., Steinberg and Gardner, 1937; Stevens and Guirao, 1967; Scharf, 1978; Hellman and Meiselman, 1993). Interestingly, a threshold-dependent difference in perceived loudness was also evident even within the two categories with audiometric thresholds typically considered to be within the normal range (i.e., –10 to 0 and 5 to 15 dB HL), with individuals having higher thresholds requiring higher input levels for equivalent loudness judgments. This finding has little clinical significance, as intervention would not be considered for individuals with thresholds within these ranges; however, it does demonstrate that loudness is a graded function of threshold, even within the normal range of hearing. This finding is consistent with the observation that distortion-product otoacoustic emissions levels decrease as threshold increases from –5 dB HL to 20 dB HL (Dorn *et al.*, 1998). Thus, it would be incorrect to assume that all listeners with thresholds less than or equal to 20 dB HL have equivalent cochlear function.

The observation that the loudness percepts of participants with normal hearing and hearing loss differed more at low input levels than at high input levels is consistent with other data comparing responses in individuals with normal hearing and hearing loss. This pattern has been previously observed in loudness data (Steinberg and Gardner, 1937; Stevens and Guirao, 1967; Scharf, 1978; Hellman and Meiselman, 1993) and masking data (e.g., Oxenham and Plack, 1997; Nelson and Schroder, 1997; Lopez-Poveda and Johannesen, 2012; Johannesen *et al.*, 2014). The pattern is also evident in objective measures of cochlear function in humans. For example, distortion-product otoacoustic emission suppression tuning curves from listeners with normal hearing and listeners with mild-moderate hearing loss are

similar when high-level probes (i.e., the level of f_2) are presented to both groups (Gruhlke *et al.*, 2012).

Like many other tasks that require a voluntary response from the participant, CLS measurements may not be possible in infants, young children, patients with developmental disabilities, and adults with cognitive disorders. Under these circumstances, previous measurements of distortion-product otoacoustic emissions and CLS in the same participants suggest that distortion-product otoacoustic emissions measurements may be used to predict the loudness function (Thorson *et al.*, 2012; Rasetshwane *et al.*, 2013).

Following Heeren *et al.* (2013), we converted CLS data from loudness in CUs to a more standard measure of loudness level (phon). This conversion involves two steps. The first step uses data at 1 kHz and the definition of phon to derive a CU-to-phon conversion (Fig. 6). The second step views CLS as a loudness-matching procedure and uses the dependence of loudness on hearing loss (see Fig. 5) to derive SPL-to-phon conversion functions for other frequencies and for different hearing losses. As hearing loss increased, the SPL-to-phon conversion functions became steeper (Fig. 7). Comparisons of the SPL-to-phon conversion functions derived for 0, 20, 40, and 60 dB HL show that, for the same loudness level, the differences in stimulus level are largest at low levels and these differences increase as hearing loss increases, and decrease as level increases. At the lowest stimulus levels, the displacement of the SPL-to-phon conversion function for 1 kHz for the group with hearing loss is approximately equal to the hearing loss (in dB HL) for that group, consistent with the observations of Steinberg and Gardner (1940).

Our motivations for converting CLS data to loudness level in phons were to allow for the construction of ELCs and to permit comparison to previous data. In comparison to ELCs from participants with normal hearing, the vertical spacing of the ELCs decreased as magnitude of the hearing loss increased (Fig. 8). The steep slopes of the CLS functions and the reduced spacing between ELCs are manifestations of abnormally rapid growth of loudness and provide evidence that CLS data are capable of showing loudness recruitment, which can also be viewed as a demonstration of reduced cochlear compression.

ELCs constructed from CLS for participants with normal hearing are consistent with previously reported data for loudness level (e.g., ISO, 2003; Suzuki and Takeshima, 2004; Heeren *et al.*, 2013); however, differences were observed. Compared to the normal ISO contours (see Fig. 9), the current contours were higher in the low frequencies for 60 phons and above. That is, the spacing of the present contours was more uniform across frequency, while the ISO contours had narrower spacing at low frequencies. The narrower spacing of the ISO contours at low frequencies has been suggested to indicate that the auditory system is overall less compressive at low frequencies (Suzuki and Takeshima, 2004), a fact that is consistent with other experimental measures of responses of the auditory system (e.g., Delgutte, 1990; Lopez-Poveda *et al.*, 2003; Gorga *et al.*, 2011). The wider spacing in the contours at low frequencies in our measurements, as well as in other CLS-derived ELCs

(Heeren *et al.*, 2013; Allen *et al.*, 1990), suggests that CLS measurements may not be as sensitive to differences in compressive behavior when compared to traditional measures of loudness. Another difference between the present contours and the ISO contours is that loudness matching for CLS was performed after the fact by assuming equality of loudness for equal CUs. We did not use the paired-comparison technique in which listeners adjust the level of a test signal to match the loudness of a reference signal that was used to obtain the ISO data. In support of this view, it has been shown that in loudness matching, the loudness function observed when the level of the reference tone was varied differed from the function that was observed when the level of the test tone was varied (Suzuki *et al.*, 1989). Thus, although there is general agreement, there are small differences among reported ELCs from participants with normal hearing. Another limitation of our CLS-derived ELCs when compared to the ISO ELCs is sample size. Although the current study included 148 participants, the ISO ELCs included a larger sample size since they were based on data from multiple studies.

The level-dependent loudness loss (Fig. 10) provides a characterization of auditory function for suprathreshold sounds. This representation may be useful for hearing-aid gain prescription. Not surprisingly, in this representation, hearing loss for suprathreshold sounds shows the greatest amount of loss (i.e., deviation from normal) for soft sounds, and less loss for loud sounds. This indicates that, for low-level inputs, more gain would be needed to reach a normal perception of loudness. The perceived loudness for high-level sounds deviates less from normal and, as a result, less gain would be required to restore normal loudness. The notion that gain should be reduced as input level is increased is not new, and, in fact, underlies the use of wide dynamic range compression in hearing aids (e.g., Cornelisse *et al.*, 1995; Byrne *et al.*, 2001). The present approach, however, can be used to describe the gain to achieve normal loudness for essentially all input levels and it provides a framework for doing it on an individual basis. This may be important because participants with similar audiometric thresholds may have different loudness-loss functions. For example, participants A and B (Fig. 10) have the same thresholds at 0.5, 1, and 2 kHz (35, 55, and 55 dB HL, respectively), but they have loudness losses for the categories shown in this figure that differ by more than 10 dB. This suggests that, while the use of average data is informative, it may not be possible to predict loudness loss from audiometric thresholds on an individual basis, and further provides support for the use of suprathreshold measures, such as loudness, for hearing-aid fitting. Given the variability in loudness percepts for patients having the same audiometric thresholds, such individualization has the potential to lead to customized signal-processing approaches and, in turn, greater listener satisfaction with amplification.

Our level-dependent loudness loss is an extension of previous approaches for fitting hearing aids using loudness measurements (e.g., Kollmeier *et al.*, 1993; Cox, 1995; Ricketts, 1996). We extended these previous studies by

specifying level-dependent loudness loss for broad regions of dynamic range of hearing, for a wide range of frequencies, and in a large number of participants.

The loudness-loss functions for softer sounds can be less than that for louder sounds for some participants (e.g., the loudness-loss function for “medium” loudness was less than that for “very loud” loudness at 0.25 and 0.5 kHz for participant C in Fig. 10). These observations were unexpected. Additionally, the loudness-loss functions were occasionally equal to or less than zero for some participants, indicating that these participants with hearing loss perceived high-level stimuli as being louder than the perceived loudness for a listener with normal hearing for the equivalent input level. These observations may be indications of hyperacusis, i.e., sensitivity to loud sounds (e.g., Anari *et al.*, 1999; Formby *et al.*, 2003), or they may suggest that there is residual measurement error and that there is a need for improvements to our methods of estimating level-dependent loudness loss on an individual basis.

Interaural attenuation can play a role in loudness judgments if there is a significant difference in audiometric thresholds between a participant’s ears. In the current study, if both ears met our inclusion criteria for participants with hearing loss, the test ear was selected as the ear with thresholds in the mild-to-moderate range. In most cases, this was the better ear. However, because the measurements were made at multiple frequencies and in the same ear for all frequencies in each participant, there were instances where measurements were not made in the better ear. Of the total 504 measurements made in participants with hearing loss (see Table I), only two measurements, from two different participants, were made in the worse ear when the interaural difference was ≥ 40 dB (specifically, 45 dB for both measurements). Thus, the stimulus in the test ear might have crossed to the non-test ear and influenced loudness judgments for these two measurements. However, the overall impact of interaural attenuation on the findings of this study is small since these measurements represent only 0.4% of the total data.

V. CONCLUSIONS

The observed CLS test-retest reliability supports its use in the assessment of hearing status in relatively untrained listeners. The ELCs derived from CLS measurements are in good agreement with standard ELCs (ISO, 2003). The conversion of CUs to phons is recommended because it facilitates comparison of CLS measurements to traditional measures of loudness level and may therefore lead to a wider acceptance of CLS. Individual estimates of level-dependent loudness loss have the potential to provide a prescription for fitting hearing aids that is more informative than audiometric thresholds.

ACKNOWLEDGMENTS

This research was supported by the NIH-NIDCD grants R03 DC013982 (D.M.R.), R01 DC2251 (M.P.G.), R01 DC8318 (S.T.N.), and P30 DC4662.

¹In the current study the median level was calculated, as opposed to the mean level, as was done in Rasetshwane *et al.* (2013). ISO (2006) recommends using the median; however, analysis of group data showed that the mean and median (across participants) CLS functions were similar.

²The IQRs reported for ISO (2006) and Heeren *et al.* (2013) are approximations obtained by digitization of figures in these papers.

³IQR for analysis of CLS data with removal of outliers can be larger than for the analysis without removal of outliers because the IQR are based on analysis of the fitted CLS functions and not the raw data. When analyses are based on raw data, IQRs without removal of outliers are always larger or equal to IQRs with removal of outliers.

- Allen, J. B., Hall, J. L., and Jeng, P. S. (1990). “Loudness growth in 1/2-octave bands (LGOB)—A procedure for the assessment of loudness,” *J. Acoust. Soc. Am.* **88**, 745–753.
- 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**, 567–578.
- Anari, M., Axelson, A., Eliasson, A., and Magnusson, L. (1999). “Hypersensitivity to sound,” *Scand. Audiol.* **28**, 219–230.
- Anweiler, A.-C., and Verhey, J. L. (2006). “Spectral loudness summation for short and long signals as a function of level,” *J. Acoust. Soc. Am.* **119**, 2919–2928.
- Blamey, P. J., and Martin, L. F. A. (2009). “Loudness and satisfaction ratings for hearing aid users,” *J. Am. Acad. Audiol.* **20**, 272–282.
- Brand, T., and Hohmann, V. (2001). “Effect of hearing loss, center frequency, and bandwidth on the shape of loudness functions in categorical loudness scaling,” *Audiology* **40**, 92–103.
- Brand, T., and Hohmann, V. (2002). “An adaptive procedure for categorical loudness scaling,” *J. Acoust. Soc. Am.* **112**, 1597–1604.
- Byrne, D., Dillon, H., Ching, T., Katsch, R., and Keidser, G. (2001). “NAL-NL1 procedure for fitting nonlinear hearing aids: Characteristics and comparisons with other procedures,” *J. Am. Acad. Audiol.* **12**, 37–51.
- Cohen, J. (1992). “A power primer,” *Psychol. Bull.* **112**, 155–159.
- Cornelisse, L. E., Seewald, R. C., and Jamieson, D. G. (1995). “The input/output formula: A theoretical approach to the fitting of personal amplification devices,” *J. Acoust. Soc. Am.* **97**, 1854–1864.
- Cox, R. M. (1995). “Using loudness data for hearing aid selection: The IHAF approach,” *Hear. J.* **48**, 10, 39–44.
- Cox, R. M., Alexander, G. C., Taylor, I. M., and Gray, G. A. (1997). “The contour test of loudness perception,” *Ear Hear.* **18**, 388–400.
- Cronbach, L. J. (1951). “Coefficient alpha and the internal structure of tests,” *Psychometrika* **16**, 297–334.
- Delgutte, B. (1990). “Two-tone rate suppression in auditory-nerve fibers: Dependence on suppressor frequency and level,” *Hear. Res.* **49**, 225–246.
- Dorn, P. A., Piskorski, P., Keefe, D. H., Neely, S. T., and Gorga, M. P. (1998). “On the existence of an age/threshold/frequency interaction in distortion product otoacoustic emissions,” *J. Acoust. Soc. Am.* **104**, 964–971.
- Elberling, C. (1999). “Loudness scaling revisited,” *J. Am. Acad. Audiol.* **10**, 248–260.
- Florentine, M., Buus, S., and Poulsen, T. (1996). “Temporal integration of loudness as a function of level,” *J. Acoust. Soc. Am.* **99**, 1633–1644.
- Formby, C., Sherlock, L. P., and Gold, S. L. (2003). “Adaptive plasticity of loudness induced by chronic attenuation and enhancement of the acoustic background,” *J. Acoust. Soc. Am.* **114**, 55–58.
- Garnier, S., Micheyl, C., Arthaud, P., Berger-Vachon, C., and Colett, L. (1999a). “Temporal loudness integration and spectral loudness summation in normal-hearing and hearing-impaired listeners,” *Acta Otolaryngol.* **119**, 154–157.
- Garnier, S., Micheyl, C., Arthaud, P., and Colett, L. (2000). “Effect of frequency content on categorical loudness normalization,” *Scand. Audiol.* **29**, 253–269.
- Garnier, S., Micheyl, C., Berger-Vachon, C., and Colett, L. (1999b). “Effect of signal duration on categorical loudness scaling in normal and in hearing-impaired listeners,” *Audiology* **38**, 196–201.
- Gorga, M. P., Neely, S. T., Kopun, J., and Tan, H. (2011). “Distortion-product otoacoustic emission suppression tuning curves in humans,” *J. Acoust. Soc. Am.* **129**, 817–827.
- Gruhlke, A., Birkholz, C., Neely, S. T., Kopun, J., Tan, H., Jesteadt, W., Gorga, M. P., and Schmid, K. K. (2012). “Distortion-product otoacoustic emission suppression tuning curves in hearing-impaired humans,” *J. Acoust. Soc. Am.* **132**, 3292–3304.

- Heeren, W., Hohmann, V., Appell, J. E., and Verhey, J. L. (2013). "Relation between loudness in categorical units and loudness in phons and sones," *J. Acoust. Soc. Am.* **133**, EL314–EL319.
- Hellman, R. P. (1999). "Cross-modality matching: A tool for measuring loudness in sensorineural impairment," *Ear Hear.* **20**, 193–213.
- Hellman, R. P., and Meiselman, C. H. (1993). "Rate of loudness growth for pure tones in normal and impaired hearing," *J. Acoust. Soc. Am.* **93**, 966–975.
- ISO (2003). 226:2003, "Acoustics-normal equal-loudness-level contours" (International Organization for Standardization, Geneva, Switzerland).
- ISO (2006). 16832:2006, "Acoustics-loudness scaling by means of categories" (International Organization for Standardization, Geneva, Switzerland).
- Johannesen, P. T., Pérez-González, P., and Lopez-Poveda E. A. (2014). "Across-frequency behavioral estimates of the contribution of inner and outer hair cell dysfunction to individualized audiometric loss," *Front Neurosci.* **8**, 214.
- Kollmeier, B., Peissig, J., and Hohmann, V. (1993). "Real-time multiband dynamic compression and noise reduction for binaural hearing aids," *J. Rehabil. Res. Dev.* **30**, 82–94.
- Leibold, L. J., Tan, H., Khaddam, S., and Jesteadt, W. (2007). "Contributions of individual components to the overall loudness of a multitone complex," *J. Acoust. Soc. Am.* **121**, 2822–2831.
- Lieberman, M. C., and Dodds, L. W. (1984). "Single-neuron labeling and chronic cochlear pathology. III. Stereocilia damage and alterations of threshold tuning curves," *Hearing Res.* **16**, 55–74.
- Lopez-Poveda, E. A., and Johannesen, P. T. (2012). "Behavioral estimates of the contribution of inner and outer hair cell dysfunction to individualize audiometric loss," *J. Assoc. Res. Otolaryngol.* **13**, 485–504.
- Lopez-Poveda, E. A., Plack, C. J., and Meddis, R. (2003). "Cochlear nonlinearity between 500 and 8000 Hz in normal-hearing listeners," *J. Acoust. Soc. Am.* **113**, 951–960.
- Marks, L. E., and Florentine, M. (2011). "Measurement of loudness, part I: Methods, problems, and pitfalls," in *Loudness*, edited by M. Florentine, A. N. Popper, and R. R. Fay (Springer, New York), pp. 17–56.
- Moore, B. C. J., Glasberg, B. R., and Stone, M. A. (1999). "Use of a loudness model for hearing aid fitting. III. A general method for deriving initial fittings for hearing aids with multi-channel compression," *Br. J. Audiol.* **33**, 241–258.
- Nelson, D. A., and Schroder, A. C. (1997). "Linearized response growth inferred from growth-of-masking slopes in ears with cochlear hearing loss," *J. Acoust. Soc. Am.* **101**, 2186–2201.
- Oetting, D., Brand, T., and Ewert, S. D. (2014). "Optimized loudness-function estimation for categorical loudness scaling data," *Hearing Res.* **316**, 16–27.
- Oxenham, A. J., and Plack, C. J. (1997). "A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing," *J. Acoust. Soc. Am.* **101**, 3666–3675.
- Rasetshwane, D. M., Neely, S. T., Kopun, J. G., and Gorga, M. P. (2013). "Relation of distortion-product otoacoustic emission input-output functions to loudness," *J. Acoust. Soc. Am.* **134**, 369–383.
- Rasmussen, A. N., Olsen, S. O., Borgkvist, B. V., and Nielsen, L. H. (1998). "Long-term test-retest reliability of category loudness scaling in normal-hearing subjects using pure-tone stimuli," *Scand. Audiol.* **27**, 161–167.
- Ricketts, T. A. (1996). "Fitting hearing aids to individual loudness-perception measures," *Ear Hear.* **17**, 124–132.
- Ricketts, T. A., and Bentler, R. A. (1996). "The effect of test signal type and bandwidth on the categorical scaling of loudness," *J. Acoust. Soc. Am.* **99**, 2281–2287.
- Robinson, K., and Gatehouse, S. (1996). "Test-retest reliability of loudness scaling," *Ear Hear.* **17**, 120–123.
- Scharf, B. (1978). "Loudness," in *Handbook of Perception: Vol. IV. Hearing*, edited by E. C. Carterette and M. P. Friedman (Academic Press, New York), pp. 187–242.
- Shaw, E. A. G. (1965). "Earcanal pressure generated by a free sound field," *J. Acoust. Soc. Am.* **39**, 465–470.
- Shi, L.-F., Doherty, K. A., and Zwislocki, J. J. (2007). "Aided loudness growth and satisfaction with everyday loudness perception in compression hearing aid users," *J. Am. Acad. Audiol.* **18**, 206–219.
- Steinberg, J. C., and Gardner, M. B. (1937). "The dependence of hearing impairment on sound intensity," *J. Acoust. Soc. Am.* **9**, 11–23.
- Steinberg, J. C., and Gardner, M. B. (1940). "On the auditory significance of the term hearing loss," *J. Acoust. Soc. Am.* **11**, 270–277.
- Stevens, S. S. (1946). "On the theory of scales of measurement," *Science* **103**, 677–680.
- Stevens, S. S., and Guirao, M. (1967). "Loudness functions under inhibition," *Percept. Psychophys.* **2**, 459–465.
- Suzuki, S., Suzuki, Y., Kono, S., Sone, T., Kumagai, M., Miura, H., and Kado, H. (1989). "Equal-loudness level contours for pure tone under free field listening condition (I)—Some data and considerations on experimental conditions," *J. Acoust. Soc. Jpn. E* **10**, 329–338.
- Suzuki, Y., and Takeshima, H. (2004). "Equal-loudness contours for pure tones," *J. Acoust. Soc. Am.* **116**, 918–933.
- Thorson, M. J., Kopun, J. G., Neely, S. T., Tan, H., and Gorga, M. P. (2012). "Reliability of distortion-product otoacoustic emissions and their relation to loudness," *J. Acoust. Soc. Am.* **131**, 1282–1289.
- Wagner, W., Heppelmann, G., Vonthein, R., and Zenner, H. P. (2008). "Test-retest repeatability of DPOAEs," *Ear Hear.* **29**, 378–391.
- Zwislocki, J. J. (1969). "Temporal summation of loudness: An analysis," *J. Acoust. Soc. Am.* **46**, 431–441.