



Published in final edited form as:

Ear Hear. 2010 August ; 31(4): 567–578. doi:10.1097/AUD.0b013e3181da4d15.

Reliability of Categorical Loudness Scaling and its Relation to Threshold

Sarah C. Al-Salim^{1,2}, Judy G. Kopun², Stephen T. Neely², Walt Jesteadt², Bettina Stiegemann², and Michael P. Gorga²

¹University of Nebraska, Lincoln, NE

²Boys Town National Research Hospital, Omaha, NE

Abstract

Objective—To further examine the reliability of categorical loudness scaling (CLS) for individual loudness categories and for the slope of the CLS functions. In addition, the relationship between CLS and audiometric threshold will be evaluated.

Design—CLS functions were obtained on 74 subjects, 58 with hearing loss and 16 with normal hearing. CLS functions were measured at three frequencies (1, 2, and 4 kHz) in two separate sessions separated by as little as 1 week and as much as 6 months. Reliability of mean and median levels within each loudness category was assessed using standard deviations and correlation coefficients. Lines were fit to the CLS functions and slopes of the lines were used to assess reliability and the relation between CLS and audiometric threshold.

Results—Similar reliability for CLS measurements was observed in both normal-hearing and hearing-impaired subjects at all frequencies. Across both groups of subjects, correlations describing the reliability of mean stimulus-level within category exceeded 0.92 at all frequencies. In addition, standard deviations of the mean stimulus-level difference between visits ranged from 6.6 to 7.8 dB, depending on frequency. The correlation between the slope of a straight line fitted to the entire CLS function and audiometric threshold collapsed across frequencies was 0.72. A segmented line was then fit to the CLS function: one segment was fit to the soft portion of the CLS function (categorical units, CU, < 20) and the other segment to the loud portion (CU > 20). Slopes of the line fit to the entire CLS function and of the line fit to the soft portion of the CLS function were both reliable across sessions. The slope of the line fit to the soft portion increased as audiometric threshold increased, with the correlations greater than 0.86 at all frequencies. No relationship was observed between slope of the line fit to the loud portion of the CLS function and audiometric threshold. Iso-loudness contours were constructed from the CLS data and used to determine the gain that would be needed to produce “normal” loudness percepts for hearing-impaired individuals.

Conclusions—Within-subject CLS measurements were reliable across sessions both for individual loudness categories and for slope of the CLS functions. In addition, the slope of the low-level portion of the CLS function varied in a predictable manner with audiometric threshold, with slope increasing as audiometric threshold increased. Finally, gain as a function of input level needed to provide loudness percepts for individuals with hearing loss equal to the loudness percepts of normal-hearing individuals can be estimated from audiometric threshold. This finding supports the assumption that both audiometric threshold and response growth (loudness) are determined by the same underlying cochlear mechanisms.

INTRODUCTION

Loudness is defined as a listener's subjective response to the intensity of a sound (Scharf, 1978). A number of methods including loudness matching, loudness balancing, magnitude estimation, magnitude production, and categorical loudness scaling have been used to measure an individual's perception and growth of loudness (see Launer, 1995, for a review). Many of these loudness-measurement procedures require training and time commitments that would be impractical for routine use in the clinic. In addition, many of these procedures do not provide direct information regarding human perception in a way that can be used clinically because they lack descriptors relating to perception. In the absence of such descriptors, it is also likely that the normal relationship between loudness categories could not be maintained. The exception to this is categorical loudness scaling (CLS), where signals of different intensities are presented to subjects who are asked to rate the loudness of the signals using meaningful labels (e.g. soft, loud, and comfortable). This method can be used to assess loudness over the complete auditory dynamic range in a matter of minutes, thus making the procedure potentially useful in the clinic. Not only is the task more easily learned and understood by the subject, the category labels may provide a clinician with information that can be used to maximize audibility and avoid loudness discomfort in hearing-aid fittings. CLS also relates more to a listener's experience and informal descriptions of their percepts of loudness.

The data obtained from CLS typically are plotted in the form of a loudness function that monotonically relates loudness perception to stimulus intensity across a listener's entire dynamic range from too soft to too loud (Jenstad et al., 1997). Features of the CLS loudness function are determined, at least in part, by the individual's audiometric thresholds. For example, the loudness functions observed in normal-hearing listeners for pure-tone stimuli typically consist of a shallow portion from near threshold to moderate levels and a steep portion at higher levels. The length and slope of the individual portions may vary depending on other factors, even among listeners with similar thresholds. Individuals with sensorineural hearing loss may provide CLS functions that include a low-level segment that begins at higher levels (related to threshold elevation) or may lack the low-level segment entirely, creating a steeply sloping high-level function, a pattern that would be consistent with an abnormally rapid growth of loudness. While there are features of CLS that make it attractive, response variability in both normal-hearing and hearing-impaired subjects may limit the clinical applicability of the measurements. The variability of CLS has not been well documented.

Loudness recruitment, observed in patients with hearing loss, has been attributed to the spread of excitation as intensity increases (Scharf, 1978), to the loss of compressive nonlinearity, which typically accompanies hearing loss of cochlear origin (Moore, 2003), or perhaps to a combination of these and/or other effects determined at more central levels of the auditory system (Heinz et al., 2005; Cai et al., 2009). Both spread of excitation and loss of compression, due to changes in basilar-membrane mechanical responses associated with damage to the outer hair cells (OHC), have been implicated as underlying causes of abnormal response growth, whereas the data reported by Heinz et al. (2005) and Cai et al. (2009) suggest an alternative (perhaps more central) explanation.

Regardless of where in the auditory system abnormal response growth is mediated, there is little doubt that loudness recruitment often accompanies the most common form of permanent sensory hearing loss, which typically begins with OHC damage. Loss of sensitivity, compression, and sharp tuning are observed when OHC damage occurs in lower animals (e.g., Dallos and Harris, 1978; Liberman and Dodds, 1987; Robles and Ruggero, 2001). It is reasonable to assume, therefore, that the initiating factor for loudness recruitment

in humans is OHC damage (Zhang and Zwislocki, 1995). Individuals with hearing losses of cochlear origin frequently present with a loss of sensitivity (reducing or eliminating the low-level portion of their loudness functions), but experience uncomfortable loudness at the same levels at which subjects with normal hearing experience uncomfortable loudness. If threshold is elevated but the level at which uncomfortable loudness occurs does not shift, then, by definition, patients with hearing loss have a reduced dynamic range and, therefore, an abnormal growth of loudness. This pattern is expected to hold for cases in which only OHC damage exists (i.e., losses of 60 dB or less). For behavioral thresholds greater than 60 dB HL, it is likely that most OHCs are damaged and no longer functioning. As hearing loss progresses beyond this point, the slope of the loudness functions does not continue to increase. At this point, any further hearing loss would be the result of damage to the inner hair cells which manifests itself in simple linear attenuation (Oxenham and Bacon, 2003).

The relationship between hearing loss and CLS functions has been described previously (i.e. Launer, 1995; Elberling, 1999). Launer (1995) fit a single linear function to the CLS data obtained from subjects with sensorineural hearing loss. On average, as behavioral threshold increased, so did the slope of the line fitted to the loudness data. However, inter-subject variability was such that, while there was a significant relationship between threshold and slope of the CLS function, it would not be possible to predict the CLS function from sensitivity loss in individual subjects. Elberling (1999) analyzed data from four studies (Elberling and Nielsen, 1993; Keissling, 1995; Launer et al., 1996; Ricketts and Bentler, 1996) in which CLS measurements were performed. The slopes of the loudness functions from all four studies were normalized relative to normal hearing and plotted with inverse slope as a function of audiometric threshold. The slope of the loudness function could be predicted from behavioral threshold in 70% of hearing-impaired listeners up to a severe degree of hearing loss, but beyond that point the dynamic range decreased at a different rate, causing the inverse slope to no longer vary linearly with hearing threshold. However, Elberling suggested that the amount of information gained from performing CLS would not justify its use clinically. Interestingly, both Launer (1995) and Elberling (1999) investigated the relationship between hearing loss and the CLS functions in terms of a single line fit to the CLS function, despite the fact that, at least for normal-hearing subjects, loudness-growth functions are often characterized by two distinct segments. The extent to which the choice of the line fit contributed to the apparent variability is undetermined.

In all of these descriptions of the CLS function, it is assumed that listeners can reliably perform the task. Indeed, CLS reliability within a single session has been described in several reports (Beattie et al., 1997; Ellis and Wynne, 1999). Test-retest differences between trials in a single session were not significant for either adults or children (Ellis and Wynne, 1999). Beattie et al. (1997) reported test-retest values within a given category of ± 12 dB 95% of the time in a single session. Both studies suggest that, within a single test session, subjects are consistent in their judgments of loudness.

Previous descriptions of test-retest reliability over multiple sessions have focused on variability within single categories, as opposed to describing within-subject variability for the entire CLS function. Some previous within-category estimates have used mean levels (Allen et al., 1990; Rasmussen et al., 1998; Ellis and Wynne, 1999; Keidser et al., 1999), while other studies used median level within each loudness category for evaluating CLS reliability, presumably on the assumption that this approach would reduce the influence of outliers (Cox et al., 1997; Palmer and Lindley, 1998). Regardless of whether mean or median levels were used, reported standard deviations were in the range of 1.4 to 9 dB with test-retest differences of less than 10 dB for adults and greater than 10 dB for children (Ellis and Wynne, 1999). Ricketts and Bentler (1996) examined variability using both mean and median data. Test-retest reliability varied depending on whether mean or median values

were used to make the calculations and on the frequency of the signal, but, regardless of the conditions, correlations greater than 0.97 were observed. In all of these studies, only small numbers of subjects were tested and few studies included hearing-impaired subjects. Data from normal-hearing individuals are needed in order to provide a reference measurement to which the loudness functions from subjects with hearing loss can be compared. However, the real interest lies in using CLS in individuals with hearing loss because the intent is to use these procedures in the clinic with patients who may experience loudness recruitment. Thus, additional data on a larger sample of both normal-hearing and hearing-impaired subjects are needed to assess the reliability of CLS measurements before they can be applied in clinical situations to describe loudness growth.

The first goal of the present study was to further assess the reliability of CLS. Differences in the sound pressure level (SPL) assigned to loudness categories and in the rate of loudness growth between two sessions were used to assess reliability. The second goal was to evaluate the relationship between behavioral threshold and loudness growth. Finally, the CLS data were used to estimate the gain required to restore normal loudness in hearing-impaired ears.

METHODS

A. Subjects

Seventy-four subjects with ages ranging from 11 to 76 years participated in this study. Sixteen of these subjects had normal hearing while 58 subjects had hearing loss. The hearing-impaired subjects had audiometric thresholds no greater than 75 dB HL. The CLS measurements were made monaurally during two sessions separated by as little as 1 week to as much as 6 months. Pure-tone thresholds were measured behaviorally using routine clinical procedures. These thresholds were measured at octave and interoctave frequencies from 0.25 to 8 kHz via air conduction and at octave frequencies from 0.25 to 4 kHz via bone conduction. The air-conduction threshold at a given frequency was classified as normal if it was less than or equal to 15 dB HL. Any audiometric threshold greater than 15 dB HL was considered to be hearing impaired. Potential subjects with an air-bone gap greater than 10 dB were excluded from the study. A 226-Hz tympanogram was measured on each subject prior to each session. Normal middle-ear status was defined as a static compliance of 0.3–2.5 cm³ and middle-ear pressure within the range of +25 to –75 daPa. Ears not meeting these criteria were excluded from further study.

B. Stimuli

The stimuli were 1000-ms pure-tones with 20-ms rise/fall times, and were separated with a 2000-ms inter-stimulus interval. Pure-tones were selected because of their established history in measures of loudness, dating back to Fletcher and Munson (1933). The pure tones were presented at the same frequencies for both threshold and CLS measurements to minimize the effects of the fine structure of hearing threshold on the interpretation of data. The tone was presented only once for each judgment. Stimuli used during measurements of categorical loudness were generated using custom-designed software [Behavioral Auditory Research Tests (BART), v2.3.32] that controlled a 24-bit sound card (CardDeluxe, Digital Audio Labs) housed in a PC and routed through a headphone buffer (HB7, Tucker-Davis Technologies) to a probe-microphone system (ER10-B±, Etymotic Research). Stimuli were presented to the subjects' ears with insert earphones (TIP-300, Nicolet). An in-the-ear calibration was performed prior to CLS measurements at each frequency. This calibration was also performed after testing at each frequency was completed to ensure that the insert-earphone placement was stable throughout the test session. In-the-ear SPL calibration was chosen as the calibration method in an effort to be consistent with other related work,

despite the evidence that alternative calibration methods (sound intensity level, forward pressure level) may avoid some of the issues associated with SPL calibrations (Siegel, 1994; Siegel and Hirohata, 1994; Neely and Gorga, 1998; Scheperle et al., 2008). Such errors, if they occurred, would be expected to cause a shift in the CLS function along the x-axis (stimulus level, dB SPL), but estimates of the slopes of the loudness function should be independent of any calibration errors.

C. Data Collection

The CLS procedures used in the current study closely followed the recommendations presented by ISO 16832 (Kinkel, 2007) for CLS measures, with the exception of several relative category names on the rating scale and slight differences in the mapping of the numerical scale.

Data were collected in a sound-treated room with the subject seated at a desk within easy reach of a touch-screen monitor on which the categories were displayed and on which the response was registered. The response scale consisted of 11 colored horizontal bars, with the length increasing as the subjective loudness category increased, as shown in Fig. 1. The number of response categories available to the listener were chosen to ensure testing ease and accuracy while avoiding limitations on the listener (Brand, 2000; Hellman, 1999). There were 6 labeled bars or categories, including “Can’t Hear, Soft, Medium, Loud, Very Loud and Too Loud” alternating with 5 unlabeled bars. Each bar also included one of a set of arbitrary numbers from 0 to 50 in steps of 5 that were used to represent categorical units (CU). The CU values acted as numerical correlates to the verbal categories and were used to facilitate data storage and analysis. A green response button cued the subjects to respond following each stimulus presentation. The next stimulus was not presented until a response was registered for the current stimulus.

Prior to data collection, subjects were instructed on the task, following an approach described by Hawkins (1980), which was modified to fit our procedures and equipment and to be appropriate for subjects 11 years of age and older. Subjects indicated their loudness judgments using the scale on the touch-screen monitor immediately after each stimulus was presented. Subjects were told to make their loudness judgment based on that presentation alone and not to compare it to previous tones that were presented in earlier conditions.

Complete CLS functions were measured at 1, 2 and 4 kHz following a training run in which 0.5 kHz was used. The starting level was set close to the midpoint between each subject’s audiometric threshold and the maximum level in an attempt to have the starting level at a similar loudness level in all subjects regardless of hearing threshold (Brand and Hohmann, 2002). For frequencies at which the threshold was within normal limits, the starting level was set at 60 dB SPL.

Similar to the reference method recommended by ISO (Kinkel, 2007), there were three phases in determining the loudness function at each frequency. Minimum and maximum levels were determined in the first phase and these levels were used as anchor points for all following trials for a given frequency at each session. The minimum level was determined to prevent presentations of stimuli at levels below the subjects’ thresholds. The minimum level was defined as the first positive judgment (CU > 0) at which the subject could hear the stimulus after responding “Can’t Hear” to a previous (lower) stimulus level. The maximum level was used to ensure levels that were uncomfortably loud were not presented to the subject. The maximum level was defined in one of two ways. In the first option, the maximum presentation level was set equal to the stimulus level just below the one in which a subject assigned a loudness of 50 CU (“Too Loud”). However, some subjects assigned a loudness of less than 50 CU to all presentations of the stimulus including the maximum

allowable level of 105 dB SPL. When this occurred, 105 dB SPL was designated as the maximum level.

The next phase consisted of recording subjects' judgments to levels between the minimum and maximum points in order to measure a complete loudness function. Four levels equidistant between minimum and maximum were presented randomly followed by five additional levels placed between the levels that had already been presented. If any two consecutive levels were judged to be more than 10 CU apart, another presentation level was added to better define the shape of the function. In the final phase, two additional trials were presented using the previously determined test levels. Thus, there were three repeated trials for each of the three test frequencies at each session with approximately four minutes required to complete a single trial, and each session was repeated once on another day. On average there were 36 stimulus presentations required to construct one loudness function from the three trials. The loudness judgments from all three trials were transformed into CUs and stored in data files.

D. Data Analysis

Signal level (dB SPL) and loudness data (CU) from the three trials for each frequency per session were used in the analyses. At each frequency, there were three presentations of each stimulus level whose loudness was judged by the subject. Loudness is expected to grow monotonically so a high-level sound would be expected to have a higher CLS judgment than a lower-level sound. In the current study, loudness judgments were compared to other judgments at the same stimulus level and then compared to the assigned categories for adjoining stimulus levels in order to identify outliers. Within a specific stimulus level, data points that deviated by 3 or more categories from the median were excluded. In addition, data points at least 2 categories above the minimum categorical value at the adjoining upper stimulus level or at least 2 categories below the maximum categorical value at the adjoining lower stimulus level were excluded (Muller and Janssen, 2004). These rules excluded outliers in the data that would increase the likelihood of non-monotonic CLS functions.

Subjects often assigned the same loudness category to different stimulus levels within a trial. The mean and median levels (specified in dB SPL) were first determined within each loudness category per trial and then were calculated across the three trials at each frequency at each session. The mean value of the difference in the stimulus levels between the two sessions and the standard deviation of the difference for each frequency were calculated. The reliability of the mean/median values within each category across the two sessions was then quantified using correlation analyses, with the goal of determining whether the use of means or medians resulted in more repeatable measurements.

Several models have been fitted to loudness data in an attempt to describe the CLS function. Launer (1995) applied a simple linear fit to loudness data that were obtained under similar conditions to those used in the present study. Brand (2000) and Brand and Hohmann (2001) demonstrated that fitting two straight lines connected in the region of medium loudness provided a better representation of the loudness function. A complication when fitting two lines to the loudness data is determining the point of intersection between the two lines. Still, it is apparent, at least for subjects with normal hearing, that the CLS function in many cases has a shallow portion at low levels followed by a steep slope at higher levels. In the present study, separate analyses were performed in which a straight line was fit to the entire loudness function and then a straight line was fit to the shallow and steep portions of the loudness function separately. An objective approach (described below) was used to determine the breakpoint between the two lines. The slope of the lines fitted to the shallow portions of the loudness function, in addition to the slope of the simple straight-line fits to the data, was used to assess the reliability of the loudness measurements across the two

sessions, in addition to the simpler correlation analyses described above. Finally, the relationship between these parameters and audiometric threshold was assessed in order to determine the extent to which loudness growth can be predicted from audiometric threshold.

RESULTS & DISCUSSION

A. Within-subject Reliability of CLS

The first goal of this study was to describe the reliability of CLS measurements. To this end, analyses were completed on the CLS data from two sessions for all 74 subjects. There were a total of 15,985 loudness judgments for the 74 subjects, only 44 of which were judged as outliers according to the outlier-exclusion rule described above. The within-subject reliability was determined by comparing both mean and median stimulus levels to which subjects assigned each category from the two separate test sessions. These comparisons were performed once with outliers included and again with outliers excluded. The reliability of the measurements was assessed by evaluating the average stimulus-level differences, standard deviations of the average stimulus-level differences, and correlation coefficients, all of which are shown in Table 1 for each of the three test frequencies. No significant differences were observed in the reliability results when means or median were used. Given the high correlation observed between mean and median levels ($r = 0.998$) and the smaller average stimulus-level differences and standard deviations of these differences for the mean levels, all further analyses were completed using mean stimulus level within each category with outliers removed.

The average stimulus-level differences, which represent the difference between the mean SPL assigned to a category in session 1 and the mean SPL for the same category in session 2, were less than 1 dB at all three test frequencies and the standard deviations of the average stimulus-level differences ranged from 6.6 dB to 7.8 dB. The correlations between the mean stimulus level within each category from session 1 and the mean stimulus level for the same category from session 2 were used as another way to determine the reliability of the CLS measurements. These correlations are provided as insets in Fig. 2a and correspond to the results shown in the top row of Table 1. Each data point represents the mean level for each individual category used by each subject during CLS measurements. Data from normal-hearing and hearing-impaired subjects are combined for each frequency because the correlation between the mean levels from the two test sessions were similar for normal-hearing and hearing-impaired subjects, and were greater than 0.92 in every case. However, there were rare instances in which large differences were observed between the two sessions (for example, the data points at 4 kHz for which very different stimulus levels were assigned the same category by two different subjects).

The reliability of CLS measurements was also assessed by comparing the slopes of the CLS functions from session 1 and 2. For this analysis, a simple regression line was fit to each subject's entire loudness function for each frequency and session, using all categories. The slopes of the CLS functions from session 1 are plotted as a function of the slopes of the CLS functions from session 2 in Fig. 2b, with each data point representing one subject. The lowest correlation of the slopes between sessions was observed at 4 kHz ($r = 0.80$), with higher correlations at 1 and 2 kHz ($r = 0.91$ and $r = 0.94$, respectively). The correlations between slopes are lower than they were when each category was considered individually. Despite being lower, significant correlations were still observed at all frequencies, including 4 kHz. In total, the data summarized in Table 1 and in Fig. 2 suggest that CLS measurements are reliable.

B. Relations between CLS and Auditory Sensitivity

Given the reliability of the CLS measurements, we were next interested in determining the relationship between these data and auditory sensitivity. To accomplish this, we first constructed an average loudness function for each subject and for each frequency by averaging the mean stimulus levels within a category across sessions. A straight line was then fitted to the mean data to obtain a slope of the overall function for each subject. These slopes are plotted as a function of audiometric threshold (collapsed across frequency) in Fig. 3. The correlation here ($r = 0.72$) is similar to that observed by Launer (1995) ($r = 0.65$), who used similar CLS methodology and who also estimated slopes for the entire CLS function by fitting all of the data with a single straight line. When slope was plotted in log units as a function of audiometric threshold for each frequency individually, correlations ranged from a low of 0.68 (1 kHz) to a high of 0.77 (4 kHz). Plotting slope on a log scale reduced the vertical spread of the slope values for the higher audiometric thresholds. In total, these results indicate that there is a relation between the slope of the entire CLS function and audiometric threshold, but based on these analyses, the relation only accounts for about 50% of the variance.

The above slopes were based on fitting all of the CLS data for a given frequency and subject with a single line. CLS functions with distinct regions are typically observed in subjects with normal hearing and in some subjects with mild-moderate hearing loss. Fitting a single line to these data may not provide the best description of loudness growth with stimulus level. In fact, previous studies have fit two separate lines to CLS data to accommodate the change in slope (e.g., Brand, 2000; Brand and Hohmann, 2001).

One issue in performing two-line fits to the data relates to the determination of the breakpoint between the two lines. Brand and Hohmann (2001) chose 25 CU as the breakpoint; however, their justification behind choosing 25 CU as the breakpoint was not provided. To determine the optimum breakpoint separating the two portions of the CLS function, a novel, quantitative approach was taken in the current study. First, a line was fit to the full range of loudness categories, combined across frequency, and correlated with auditory sensitivity. Next, the highest level category was removed, lines were fit to the remaining individual CLS data, and the slopes were again correlated with auditory sensitivity. This process was continued, eliminating the highest remaining category, until only the two lowest categories remained. Figure 4 describes the results from this process, where slope is plotted as a function of auditory sensitivity. Each panel represents the slope for a different range of categorical units (CU), going from the most inclusive to the least inclusive range (upper left panel and lower right panel, respectively). The correlation between slope and auditory threshold is provided within each panel. There is a systematic increase in correlation as the range becomes less inclusive, reaching a maximum value when only CU's < 25 are included in the fit ($r = 0.89$). Further restrictions in the range of CU's fitted with the linear function decreased the correlation. We used these observations to define $CU = 20$ as the breakpoint between the low- and high-level portions of the CLS function. The segment of the CLS function that included CU's ≤ 20 represents the "soft" portion of the loudness function, while the segment that included CU's > 20 represents the "loud" portion of the same function. The breakpoint at $CU = 20$ might be viewed as an estimate of the medium loudness level.

Given the observation that the correlation between slope and threshold was greatest when the slope of the soft portion of the CLS function was used, we re-evaluated the reliability of the CLS data using only CU's ≤ 20 to calculate the slope. The correlation of the soft slopes from the two sessions was highest at 1 kHz ($r = 0.88$) and lowest at 4 kHz ($r = 0.76$). These correlations were lower than those observed when the entire set of CLS data for an individual subject were fitted with a straight line. It is unclear why the soft-slope

correlations are lower, but may be related to the reduced overall range in the slopes when the fits are restricted to CU's of 20 or less. Still, the correlations between the soft slopes from sessions 1 and 2 suggest that these slopes were reproducible within subjects. On the other hand, the slopes of the lines fitted to the loud portion of the CLS function were not reliable across sessions. The correlations between the loud slope from sessions 1 and 2 were -0.25 at 1 kHz, 0.04 at 2 kHz, and 0.16 at 4 kHz.

Figure 5 shows the relation between the slopes of the low-level (“soft”) and high-level (“loud”) portions of the CLS function and audiometric threshold, with circles representing soft slope and triangles representing loud slope. The correlations between soft slope and audiometric threshold are similar across frequencies, ranging from 0.86 at 1 kHz to 0.89 at 4 kHz. These correlations are higher than those seen between the slopes of straight-line fits to the entire CLS function and audiometric threshold, and indicate that nearly 80% of the variance in soft slope can be accounted for by threshold. In contrast, there is little or no correlation between loud slope and audiometric threshold at all frequencies ($r = -0.01$ at 1 kHz, $r = 0.28$ at 2 kHz, and $r = 0.21$ at 4 kHz). Less than 10% of the variance in loud slope can be accounted for by audiometric threshold. Loud slope could not be determined for a number of subjects because they did not use two or more categories > 20 CU so slope could not be calculated. As a result, 10 subjects at 1 kHz, 26 subjects at 2 kHz, and 22 subjects at 4 kHz were not included in the loud-slope analyses.

The relationship between soft slope and audiometric threshold allows the use of group-average linear regression coefficients (i.e. slope and intercept) to predict the loudness perception within the soft range of an individual using the following equation:

$$\text{Loudness (CU)} = \text{Intercept (CU)} + \text{Slope (CU/dB)} * \text{Input Level (dBSPL)}$$

The average slopes and average intercepts were calculated separately for each frequency and audiometric threshold and are provided in Table 2. When an individual's audiometric threshold is known, the equation can be used to determine an estimate of the loudness category (CU) that individual would assign to any input level. This equation holds true only for loudness percepts within the range of inaudible to medium loud and not for the loud range where little relation is observed between slope and audiometric threshold.

The correlations between SPL at the breakpoint (20 CU) separating the soft and loud portions of the CLS functions and audiometric threshold were low, never exceeding 0.39 . Because the breakpoint is independent of audiometric threshold (i.e., it occurs within the same range of stimulus levels regardless of threshold), it follows that soft slope would become steeper as the amount of hearing loss becomes greater because the softest level that is audible increases as threshold increases.

C. Iso-loudness Contours

Iso-loudness contours were constructed by finding the median level in dB SPL assigned to each loudness category (from CU=5 to CU=45) for each audiometric-threshold group (from -5 dB HL to 55, 60 or 75 dB HL, depending on frequency). These data are plotted in Fig. 6, with CU as the parameter. The lines represent best fits to the data for each CU. As expected, the slopes of these loudness contours decrease as CU increases, meaning that the greatest differences between normal-hearing and hearing-impaired subjects in SPL for equal loudness occur for low-level stimuli. Differences in SPL for equivalent loudness decrease as loudness increases. For loudness judgments of CU > 30 , there is little or no effect of threshold on the SPL for equivalent loudness. Loudness recruitment is also evident in the results summarized in Fig. 6. For example, compare the range of levels from just audible

(CU=5) to very loud (CU=45) in subjects with thresholds of 0 dB HL to the same range of loudness for subjects with thresholds of 60 dB HL. For normal-hearing subjects, SPL varied by 75–80 dB, whereas for the subjects with moderate hearing loss, SPL varied by only 20–40 dB for the same loudness range.

Figure 7 shows hypothetical CLS functions for one normal-hearing and one hearing-impaired subject, which are consistent with the dependence of loudness range on threshold evident in the mean data shown in Fig. 6. The dB difference between the normal-hearing and the hearing-impaired ear (i.e., the length of the horizontal arrows) might be used to determine the gain as a function of input level for the hearing-impaired subject that would be needed for this subject to experience the same loudness as the normal-hearing subject.

From the lines fit to the data in Fig. 6, we can develop a normative “reference” by determining the input level for a given loudness for subjects with normal hearing. From these same data, we can determine the input level that would be needed to elicit the same loudness judgment for subjects with hearing loss. We can then subtract the “normal reference” input level from the input level for the hearing-impaired subject to determine how much gain the hearing-impaired subject would require to experience the same loudness percept as the subject with normal hearing. Figure 8 plots this gain estimate as a function of input level, with audiometric threshold as the parameter. These estimates are equivalent to the way gain was determined from the individual data shown in Fig. 7, only here we are using functions fitted to group data to make the calculations.

The summary in Fig. 8 describes gain as a function of input level that would be needed to restore normal loudness for each of the six hearing-loss categories shown in the figure. As expected, the amount of gain decreases as input level increases, reaching 0–5 dB once the input level is about 100 dB SPL. This upper limit does not depend on audiometric threshold. These functions also indicate that gain (for normal loudness) grows nonlinearly, with the change in gain with input level more rapid as hearing loss increases. These data are not surprising, given the known changes in loudness as a consequence of hearing loss.

GENERAL DISCUSSION

The results from this study are summarized by the following observations.

- A. Within-subject CLS measurements are reliable across sessions both for individual loudness categories and for slopes of CLS functions.
- B. The slope of the low-level, soft portion of the CLS function varies systematically with audiometric threshold, with slope increasing as threshold increases.
- C. Gain as a function of input level needed to provide loudness percepts for individuals with hearing loss equal to the loudness percepts of normal-hearing individuals can be estimated from audiometric threshold.

A. Within-subject Reliability of CLS Measurements

The reliability of CLS measurements for individual loudness categories was determined using both mean and median stimulus levels within each category, once with outliers included and once with outliers excluded. As can be seen in Table 1, differences were observed in reliability depending on whether mean or median stimulus levels were used; however, the differences were small and provided outliers exerted only a minimal influence on the data. Ricketts and Bentler (1996) also examined reliability using both mean and median data and concluded that using median rather than mean levels within each category led to more reliable scores. When using median level within each loudness category,

Ricketts and Bentler reported a test-retest correlation for signals centered at 0.5 kHz of 0.99, which equated to an average stimulus-level difference between sessions of 3.3 dB. The test-retest correlation was 0.98 for signals centered at 3.15 kHz, which corresponded to an average stimulus-level difference of 4.1 dB. In comparison, when mean levels were used, the test-retest correlations were slightly lower across frequencies ($r = 0.97$), with larger stimulus-level differences of approximately 5 dB.

However, in the current study, the highest reliability was observed when the mean stimulus levels within each category, with outliers removed, were used in the analyses and consequently, for the analyses relating sensitivity loss to loudness. Subjects produced loudness judgments that were consistent across sessions, with mean stimulus-level differences across sessions of less than 1 dB. The present small mean differences between SPL for session 1 versus SPL for session 2 is a consequence of averaging values, which were sometimes greater for session 1 and sometimes greater for session 2. Maintaining the sign of these differences to calculate mean stimulus-level differences resulted in the small values in the present study. Ricketts and Bentler (1996) did not state how their differences were calculated. It is possible that they used absolute differences in their calculations, which would have resulted in larger mean stimulus-level differences, compared to the present results.

The present standard deviations of the mean stimulus-level differences of individual CLS category measurements were within the range reported by other studies that measured between-session CLS reliability similarly (Palmer and Lindley, 1998; Cox et al., 1997; Allen et al., 1990; Rasmussen et al., 1998; Keidser et al., 1999). When using mean levels within category, the average within-subject standard deviation of the test-retest difference was reported as 2.9 dB (Allen et al., 1990). Rasmussen et al. (1998) and Keidser et al. (1999) found within-subject standard deviations of the test-retest differences to range from 1.4 dB to about 9 dB. When using median levels within category, Palmer and Lindley (1998) reported differences in each category that ranged from 0.7 dB to 4.6 dB, with standard deviations from 3.5 dB to 8.1 dB across the frequencies tested. Cox et al. (1997) found most individual test-retest differences to be less than 6 dB and only 1–2% of the differences were greater than 10 dB. The standard deviations of these differences ranged from 2 to 6 dB.

In comparison, the CLS measurements in the present study were obtained from a larger number of hearing-impaired subjects ($n = 58$) than in previous studies. Correlations greater than 0.92 were observed at all frequencies regardless of whether the subjects were normal-hearing or hearing-impaired, and standard deviations were between 6.6 and 7.8 dB, depending on frequency. Thus, the results from the present study and those reported previously indicate that CLS measurements produce reliable results. Of particular interest in the present study is the observation that the reliability of the measurements was similar for normal-hearing and hearing-impaired subjects, thus supporting the use of the measurements in the clinic.

If CLS measurements are to be used clinically, then it is of interest to compare CLS reliability to the reliability of the audiometric threshold measured behaviorally in the audiology clinic. The reliability of audiometric thresholds has been investigated previously (Jervall and Arlinger, 1986; Burns and Hinchcliffe, 1957) and reported in the form of correlations and standard deviations of the dB difference between sessions. Jervall and Arlinger (1986) tested audiometric thresholds of normal-hearing and moderately hearing-impaired subjects during two sessions. The standard deviations of the mean difference between sessions were similar for both groups and ranged from 3.0 to 7.2 dB. Burns and Hinchcliffe (1957) tested pure-tone thresholds in normal-hearing and hearing-impaired subjects at two sessions spaced one week apart. Test-retest reliability was presented in terms

of correlations, which ranged from 0.77 to 0.86, in addition to mean level differences between sessions in the range of 1.0 to 2.2 dB with standard deviations of 4.2 to 7.6 dB. The reliability of audiometric thresholds can be viewed as something of a gold standard, since the audiometric threshold is perhaps the most fundamental clinical measurement. As demonstrated in the current study, CLS appears to have similar reliability.

Unlike previous studies where only reliability of the individual loudness categories was presented, the current study also investigated the reliability of the slopes of the CLS functions. It was found that the slopes of entire CLS functions and the slopes of the soft portion of the CLS functions are reliable within subjects across sessions. Unexpectedly, poorer reliability was observed for the slopes of the soft portion of the CLS function than for the slopes of the straight line fit to the entire CLS function. This could be a consequence of the reduced range of slope values that exist when the slope is estimated only from the lower portion of the CLS function. The slopes of the loud portion of the CLS functions were not reliable across sessions.

Simple procedural changes, including the use of an alternative in-the-ear calibration method and the implementation of an outlier exclusion rule during data collection, could improve the reliability of the CLS measurements. There were instances in which the mean stimulus levels for all of the categories at the second visit were either higher or lower by a similar number of dB compared to the mean stimulus levels from the first visit. It is possible that the differences in the loudness judgments between sessions were the result of differences in calibration, not differences in the subject's loudness perceptions or other causes of variability. Sound intensity level or forward pressure level calibrations may minimize errors resulting from the in-the-ear SPL calibrations used in the present study (Neely and Gorga, 1998; Scheperle et al., 2008).

In addition, the reliability of the CLS measurements may be improved by the use of an outlier-exclusion rule that is implemented while the subject is being tested, allowing an additional presentation of the signal at any level where a loudness judgment is determined to be an outlier. This would ensure that three loudness judgments are accepted at each stimulus level as opposed to the method used in the present study where the outlier-exclusion rule was implemented after data collection was completed. In this case, if a judgment was ruled an outlier, only two loudness judgments were available for analysis at that stimulus level.

Overall, it appears that CLS measurements result in reliable data for both normal-hearing and hearing-impaired subjects. This was true for the three frequencies tested in the present study, and it was true for comparisons of SPL within a category and in terms of the slopes of the CLS function. These results are encouraging in that they suggest that CLS can be used to address more interesting issues, such as the relation between sensitivity loss and loudness as estimated by the CLS procedure.

B. Relationship between CLS and Auditory Sensitivity

A relationship was observed between the slope of the line fit to the entire CLS function and audiometric threshold. The correlation in the present study ($r = 0.72$) was larger than the one observed by Launer (1995) ($r = 0.65$). Launer also fit entire CLS functions with a straight line and included subjects with greater degrees of hearing loss, which may have had an influence on the correlation as a large amount of inter-subject variability in the slope of the CLS function in Launer's data was observed in cases in which severe hearing loss was present. The same trend was observed in the present study where inter-subject variability in the slopes of the CLS functions among subjects with the same thresholds increased as audiometric threshold increased.

Elberling (1999) argues that it is useless to measure CLS functions because they are variable in normal-hearing subjects, thus making it difficult to establish a normative reference to which hearing-impaired data could be compared. Elberling reviewed data from 7 papers to make this point. However, a variety of definitions of normal hearing were used across those 7 studies. Thus, it is not surprising that there was variability in normative data. While it may be convenient to define normal hearing as thresholds less than or equal to 15 or 20 dB HL, it may not be valid to assume that an ear with a threshold of -5 dB HL performs exactly the same way as an ear with a threshold of 15 or 20 dB HL. Thus, the range of normal threshold would be expected to add variance to the data. Treating hearing loss as a quasi-continuous variable (even within the normal range) might have revealed differences.

Elberling (1999) also reported that the inverse slope of the loudness function varies linearly with hearing threshold and with a constant variance across thresholds. In most cases the slope of the function could be predicted from threshold. In this regard, we do not disagree, but it may be important to note that threshold accounted for 74% of the variance in the present data. Thus, there is a group of individuals for whom threshold would not accurately predict response growth. Elberling himself notes that in 30% of the hearing-impaired subjects in his study, threshold alone was not a good predictor. It is in these individuals that CLS measurement would be useful. Furthermore, in the Elberling study, the slope of the entire loudness function was assumed to be linear. In contrast, we found that the CLS functions are best fit with two lines, not one. Brand and Hohmann (2001) also recommended a revision of Elberling's conclusions due to his assumption of the loudness function being linear.

Despite the inter-subject variability observed in the present study and in Launer's study, recruitment was observed in many of the subjects with hearing loss in both studies. Even individual subjects with hearing loss at some frequencies and normal hearing at other frequencies showed normal CLS functions at frequencies where hearing was normal and steeper CLS functions where elevated behavioral thresholds existed. Slopes of the lines fit to the entire CLS functions ranged from shallow to steep within a group of subjects with the same degree of hearing loss, as can be seen in the present Fig. 3 (consider the results for subjects with audiometric thresholds of 55 dB HL). This variability observed in the CLS data within a threshold category may be the result of different underlying mechanisms being responsible for hearing losses of the same degree, although this can never be known in humans, for whom gold-standard anatomical studies are not possible.

It was expected that the CLS function would be better fit with a segmented line consisting of two sections, one fit to the softer, shallow portion of the CLS function and one to the louder, steeper portion. Because the shallow portion appears to be most affected by hearing loss, a relationship between its slope and audiometric threshold was expected. Indeed, the slope of the lines fit to the soft portions of the CLS functions varied with audiometric threshold in a predictable manner, producing correlations greater than 0.86 for each frequency. As audiometric threshold increased, the "soft" slope increased because the stimulus levels at which medium loudness was attained were similar across all threshold categories. As a result, the range of levels encompassing the loudness percepts from just audible to "Medium" is larger in normal-hearing individuals than the range of levels for the same loudness categories in a hearing-impaired individual. The inter-subject variability of the soft slope was similar across all threshold categories, unlike the trend of increasing inter-subject variability with increases in threshold when the entire CLS function was fit with a straight line.

The slope of the soft portion of the CLS function of normal-hearing subjects ranges from about 0.2 to 0.3, which is consistent with the slopes of the portion of the CLS functions < 25

CU in normal-hearing subjects reported by Brand (2000). However, the soft slopes for the subjects with the greatest amount of hearing loss in the current study are about 3 times steeper. Although no tests were performed in the current study to rule out the presence of retrocochlear pathology, these trends appear to be consistent with changes in response growth that one would predict as a consequence of damage to the normal nonlinear mechanism responsible for compression (see Robles and Ruggero, 2001, for a review). The changes seen in the low-level portion of the CLS function as a result of sensorineural hearing loss can be explained by the loss of gain as a consequence of damage to the OHCs, which was presumably the cause of the hearing losses for the subjects in this study. The loss of the gain from OHC damage results in a loss of compression which gives rise to an abnormally rapid growth in loudness known as recruitment. The relationship between the soft slope of the CLS function and audiometric threshold supports the idea that recruitment is primarily the result of peripheral changes in the auditory system, as opposed to being more of a central phenomenon (Heinz et al., 2005; Cai et al., 2009), if one accepts that threshold is determined by peripheral mechanisms.

In contrast to the relationship observed between the soft portion of the CLS function and audiometric threshold, little or no correlation was observed between the loud portion and audiometric threshold. The percept of loudness which grows abnormally fast in sensorineural hearing loss eventually is similar to the growth of loudness experienced by normal-hearing individuals. The point where hearing-impaired loudness coincides with normal loudness appears to be located close to a percept of medium loudness for individuals with audiometric thresholds less than approximately 60 dB HL. For stimuli judged to be medium loud or louder, hearing-impaired and normal-hearing individuals tend to assign the same categories to similar stimulus levels.

C. Equating Loudness in Normal and Impaired Ears

The iso-loudness contours constructed from the CLS functions (Fig. 6) provide information about the comfortable listening ranges of normal-hearing subjects and subjects with hearing loss up to a moderately-severe degree. As can be seen in Fig. 6, the range from threshold of audibility to “loud” for an individual with normal hearing is on the order of 80–90 dB while the same range of loudness in an individual with a 50-dB HL audiometric threshold is smaller, approximately 40–50 dB. Due to the upper limit of the stimulus level used in the present study (105 dB SPL), the entire dynamic range of many subjects, especially those with greater hearing losses, could not be determined. These individuals would have required stimulus levels greater than 105 dB SPL in order to perceive the sounds as either “Very Loud” or “Too Loud”.

As seen in Fig. 6, the stimulus level that elicits a loud description by the subjects is similar across hearing-threshold groups while the stimulus level needed to elicit a just audible perception varies from 20 dB SPL for normal-hearing individuals up to approximately 60–80 dB SPL for individuals with the highest degrees of hearing loss tested in the present study. As a result, greater amounts of gain are required at lower levels than at higher levels in order to restore normal loudness percepts to hearing-impaired individuals. However, once stimulus levels of 90 dB SPL or greater are reached, little or no gain is required to normalize the loudness percepts of the hearing-impaired individuals with losses of about 60 dB. The iso-loudness contours constructed from the CLS data allow us to determine the amount of gain needed to restore normal loudness percepts to individuals with no greater than a moderate degree of hearing loss (Fig. 8).

It is important to note that the gain necessary to produce equal-loudness percepts in normal and impaired ears would not necessarily be recommended for use in hearing-aid gain prescriptions, even though our gain estimates would be consistent with the gain prescribed

by National Acoustic Laboratory (NAL-NL1) in a number of cases (Byrne et al., 2001). The signals used in these CLS measurements utilized only pure-tones and, thus, may not represent the perception of loudness under realistic conditions in which the sound environment is more complex. The CLS functions obtained from more complex signals, whether they are a complex of tones or pure-tones in a background of ambient noise, would be different from those obtained with pure-tones due to the loudness summation that occurs across bandwidths (Cox et al., 1997; Brand and Hohmann, 2001). However, if CLS is performed with the intention of assisting with the fit of hearing aids, then stimuli with narrow spectra would be needed to assure that differences in loudness growth as a function of frequency are identified and taken into account in setting hearing-aid characteristics. Speech-shaped noise would be useful once these settings are established, but would not be helpful in setting I/O functions at the prescriptive stage because of its broad bandwidth, unless loudness growth was the same regardless of frequency.

Common clinical practice is to provide no gain for low-level inputs, and, in fact, expansion is sometimes used. In this way, ambient room noise is not amplified. Gain is only provided once the sound reaches a level of 20–30 dB SPL to avoid a situation in which the hearing-impaired patient is listening to amplified noise that ordinarily would not be perceived without amplification. Still, the results in the present study provide evidence that CLS (a procedure that uses meaningful labels and is clinically applicable) results in reliable estimates of loudness that are related to sensitivity loss. The present data also provide equal-loudness contours for individuals with a range of audiometric thresholds from normal hearing to moderate-to-severe hearing loss.

Acknowledgments

Support for this study was provided by the NIH-NIDCD (R01 DC2251, P30 DC4662). We would like to thank Sandy Estee and Heather Porter for their help during subject recruitment and data collection.

REFERENCES

- Allen JB, Hall JL, Jeng PS. Loudness growth in 1/2-octave bands (LGOB) -- A procedure for the assessment of loudness. *J Acoust Soc Am.* 1990; 88:745–753. [PubMed: 2212299]
- Beattie RC, Huynh RC, Ngo VN, et al. IHAF loudness contour test: Reliability and effects of approach mode in normal-hearing subjects. *J Am Acad Audiol.* 1997; 8:243–256. [PubMed: 9272746]
- Brand, T. Analysis and Optimization of Psychophysical Procedures in Audiology. Unpublished Ph.D. thesis. Oldenburg, Germany: Oldenburg University; 2000.
- Brand T, Hohmann V. Effect of hearing loss, centre frequency, and bandwidth on the shape of loudness functions in categorical loudness scaling. *Audiology.* 2001; 40:92–103. [PubMed: 11409767]
- Brand T, Hohmann V. An adaptive procedure for categorical loudness scaling. *J Acoust Soc Am.* 2002; 112:1597–1604. [PubMed: 12398465]
- Burns W, Hinchcliffe R. Comparison of the auditory threshold as measured by individual pure tone and by Bekesy audiometry. *J Acoust Soc Am.* 1957; 29:1274–1277.
- Byrne D, Dillon H, Ching T, et al. NAL-NL1 procedure for fitting nonlinear hearing aids: characteristics and comparisons with other procedures. *J Am Acad Audiol.* 2001; 12:37–51. [PubMed: 11214977]
- Cai S, Ma WD, Young ED. Encoding intensity in ventral cochlear nucleus following acoustic trauma: implications for loudness recruitment. *J Assoc Res Otolaryngol.* 2009; 10:5–22. [PubMed: 18855070]
- Cox RM, Alexander GC, Taylor IM, et al. The contour test of loudness perception. *Ear Hear.* 1997; 18:389–400.

- Dallos P, Harris D. Properties of auditory nerve responses in absence of outer hair cells. *J Neurophysiol.* 1978; 41:365–383. [PubMed: 650272]
- Elberling C. Loudness scaling revisited. *J Am Acad Audiol.* 1999; 10:248–260. [PubMed: 10331617]
- Elberling, C.; Nielsen, C. The dynamics of speech and the auditory dynamic range in sensorineural hearing impairment. In: Beilin, J.; Jensen, GR., editors. *Recent Developments in Hearing Instrument Technology.* Denmark: Danavox Jubilee Foundation; 1993. p. 99-134.
- Ellis MR, Wynne MK. Measurements of loudness growth in 1/2-octave bands for children and adults with normal hearing. *Am J Audiol.* 1999; 8:40–46. [PubMed: 10499118]
- Hawkins DB. Loudness discomfort levels: A clinical procedure for hearing aid evaluations. *J Speech Hear Disord.* 1980; 45:3–15. [PubMed: 7354628]
- Heinz MG, Issa JB, Young ED. Auditory-nerve rate responses are inconsistent with common hypotheses for the neural correlates of loudness recruitment. *J Assoc Res Otolaryngol.* 2005; 6:91–105. [PubMed: 15952047]
- Hellman R. Cross-modality matching: A tool for measuring loudness in sensorineural impairment. *Ear Hear.* 1999; 20:193–213. [PubMed: 10386847]
- Jenstad LM, Cornelisse LE, Seewald RC. Effects of test procedure on individual loudness functions. *Ear Hear.* 1997; 18:401–408. [PubMed: 9360863]
- Jervall L, Arlinger S. A comparison of 2-dB and 5-dB step size in pure-tone audiometry. *Scand Audiol.* 1986; 15:51–56. [PubMed: 3704542]
- Keidser G, Seymour J, Dillon H, et al. An efficient, adaptive method of measuring loudness growth functions. *Scand Audiol.* 1999; 28:3–14. [PubMed: 10207952]
- Keissling J. Loudness growth in sensorineural hearing loss—Consequences for hearing aid design and fitting. *Audiol Akust.* 1995; 34:82–89.
- Kinkel, M. The new ISO 16832 ‘Acoustics—Loudness scaling by means of categories’. 8th EFAS Congress/10th Congress of the German Society of Audiology; Heidelberg. 2007; 2007.
- Launer, S. Loudness Perception in Listeners with Sensorineural Hearing Impairment. Unpublished Ph.D. thesis. Oldenburg, Germany: Oldenburg University; 1995.
- Launer S, Holube I, Hohmann V, et al. Categorical loudness scaling in hearing-impaired listeners—can loudness growth be predicted from the audiogram? *Audiol Akust.* 1996; 4:156–163.
- Lieberman MC, Dodds LW. Acute ultrastructural changes in acoustic trauma: serial-section reconstruction of stereocilia and cuticular plates. *Hear Res.* 1987; 26:45–64. [PubMed: 3558143]
- Moore, BCJ. *An Introduction to the Psychology of Hearing.* New York, NY: Academic Press; 2003.
- Muller J, Janssen T. Similarity in loudness and distortion product otoacoustic emission input/output functions: Implications for an objective hearing aid adjustment. *J Acoust Soc Am.* 2004; 115:3081–3091. [PubMed: 15237833]
- Neely ST, Gorga MP. Comparison between intensity and pressure as measures of sound level in the ear canal. *J Acoust Soc Am.* 1998; 104:2925–2934. [PubMed: 9821338]
- Oxenham AJ, Bacon SP. Cochlear compression: Perceptual measures and implications for normal and impaired hearing. *Ear Hear.* 2003; 24:352–366. [PubMed: 14534407]
- Palmer CV, Lindley GA. Reliability of the Contour Test in a population of adults with hearing loss. *J Am Acad Audiol.* 1998; 9:209–215. [PubMed: 9644618]
- Rasmussen AN, Olsen SØ, Borgkvist BV, et al. Long-term test-retest reliability of category loudness scaling in normal-hearing subjects using pure-tone stimuli. *Scand Audiol.* 1998; 27:161–167. [PubMed: 9728776]
- Ricketts TA, Bentler RA. The effect of test signal type and bandwidth on the categorical scaling of loudness. *J Acoust Soc Am.* 1996; 99:2281–2287. [PubMed: 8730074]
- Robles L, Ruggiero MA. Mechanics of the mammalian cochlea. *Physiol Rev.* 2001; 81:1305–1352. [PubMed: 11427697]
- Scharf, B. Loudness. In: Carterette, EC.; Friedman, MP., editors. *Handbook of Perception: Vol IV. Hearing.* New York: Academic Press; 1978. p. 187-242.
- Scheperle RA, Neely ST, Kopun JG, et al. Influence of in-situ sound level calibration on distortion product otoacoustic emission variability. *J Acoust Soc Am.* 2008; 124:288–300. [PubMed: 18646977]

- Siegel JH. Ear-canal standing waves and high-frequency sound calibration using otoacoustic emission probes. *J Acoust Soc Am*. 1994; 95:2589–2597.
- Siegel JH, Hirohata ET. Sound calibration and distortion product otoacoustic emissions at high frequencies. *Hear Res*. 1994; 95:146–152. [PubMed: 7896573]
- Zhang M, Zwislocki JJ. OHC response recruitment and its correlation with loudness recruitment. *Hear Res*. 1995; 85:1–10. [PubMed: 7559163]

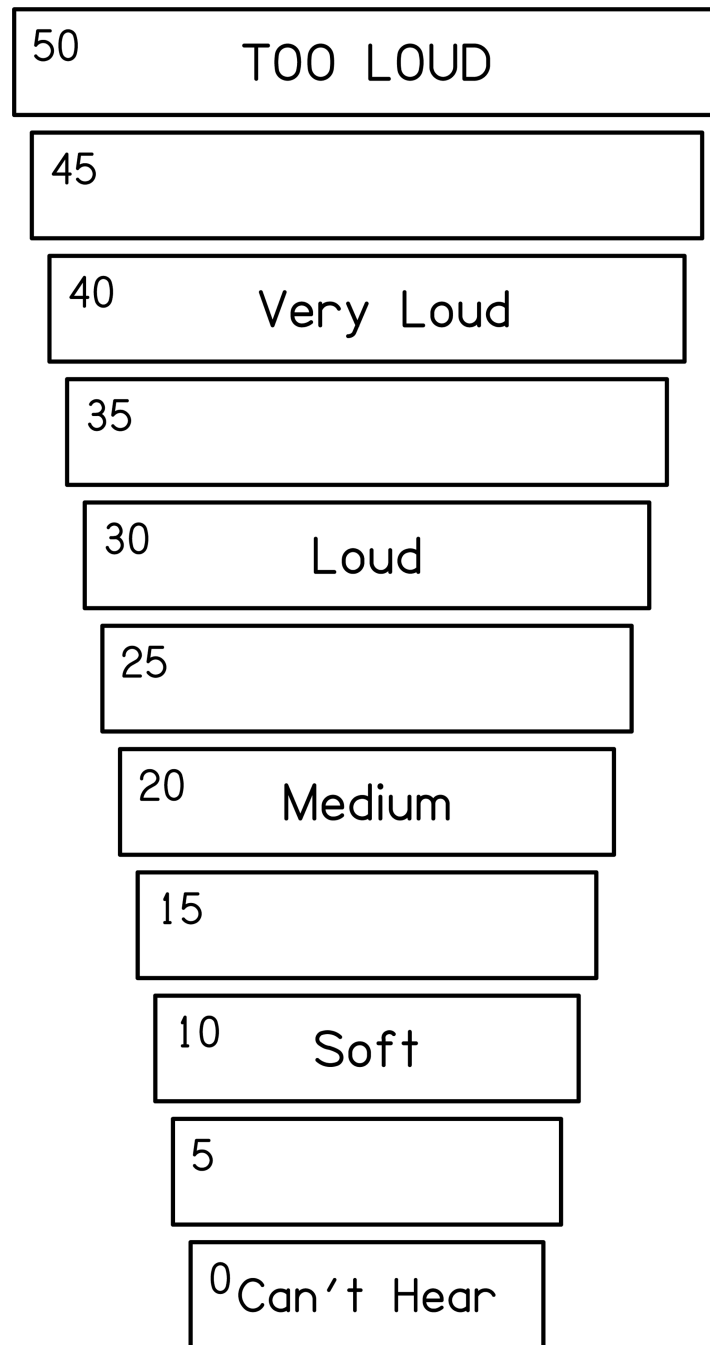
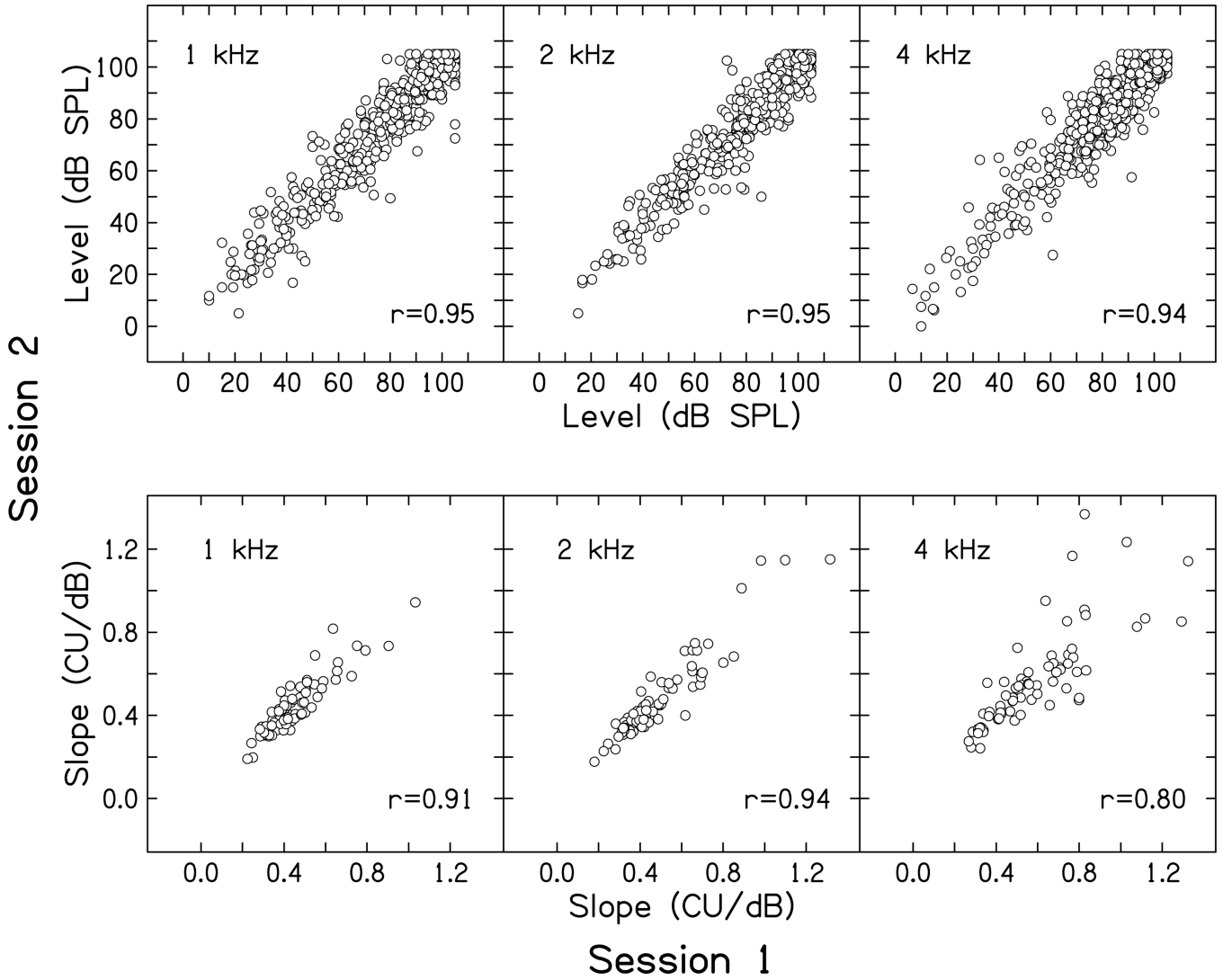


Figure 1. CLS scale with 11 response categories displayed on a touch-screen monitor for use by the subjects to rate the loudness of the signal. The numbers along the left side indicate categorical units (CU) and were used for recording the subject's responses and data analysis.



Figures 2a and 2b.

Figure 2a (top) plots the mean stimulus level from the three trials in session 2 as a function of the mean stimulus level for the three trials from session 1. Each data point represents levels for every individual category used by each subject during CLS measurements. Figure 2b (bottom) plots the slope (CU/dB) of a simple regression line fit to all of the data from each subject’s entire CLS function from session 2 as a function of the slope of the function fit to the comparable set of individual data from session 1. In both figures, data for each frequency are provided separately in each panel and correlations are provided as insets in each panel. Data from normal-hearing and hearing-impaired subjects are combined within each panel.

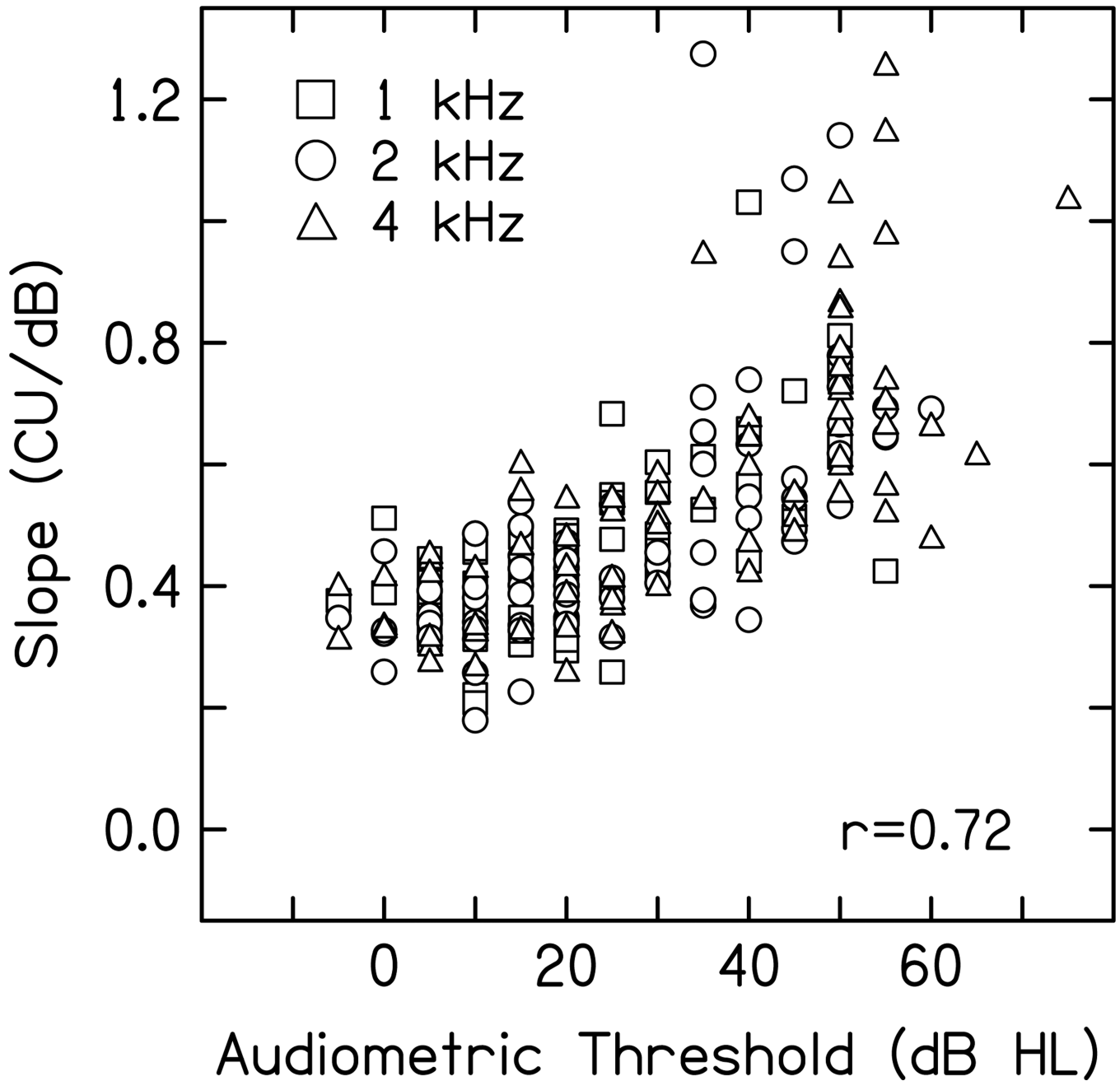


Figure 3.

Slope (CU/dB) of a straight line fit to the entire CLS function for each subject is plotted as a function of audiometric threshold, following the convention used by Launer (1995). Data for different frequencies are represented by different symbols: squares – 1 kHz; circles – 2 kHz; triangles – 4 kHz. The correlation provided within the figure was determined by collapsing across frequency.

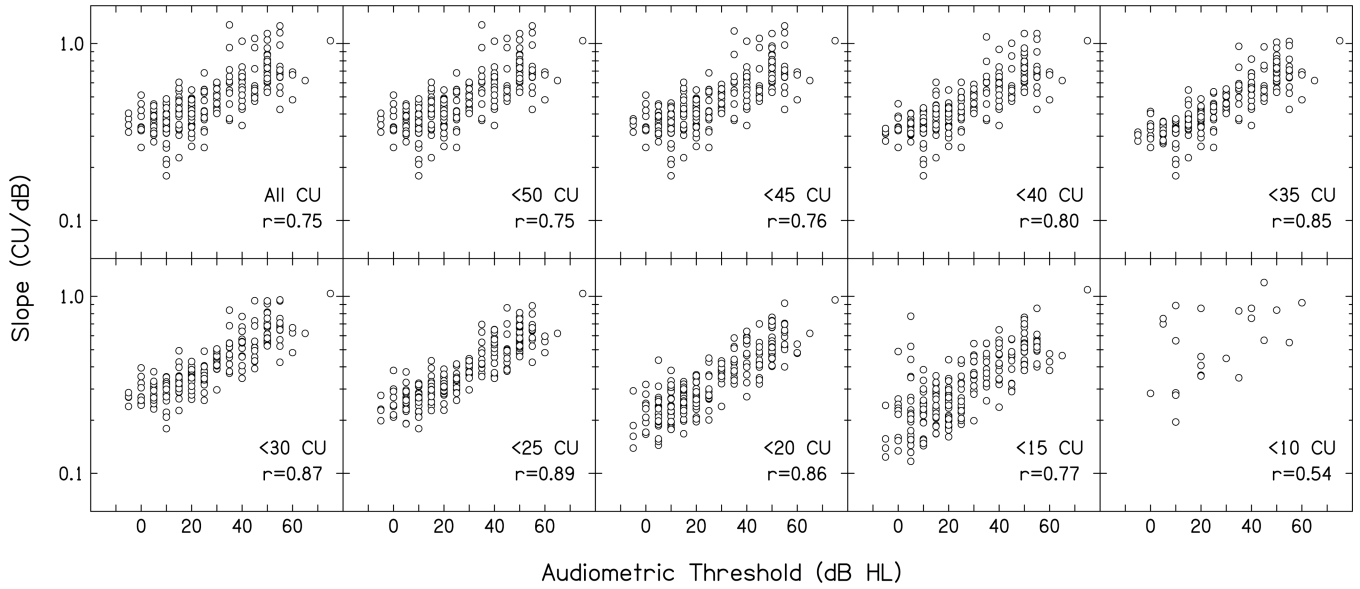


Figure 4. Slope (CU/dB) of a line fit to decreasing ranges of categories from averaged individual loudness functions are shown as a function of audiometric threshold. The slopes associated with each range of categories are plotted in separate panels, going from the most inclusive to the least inclusive range (upper left panel and lower right panel, respectively). Within each panel, data are shown for all three frequencies. Correlations are provided as insets within each panel.

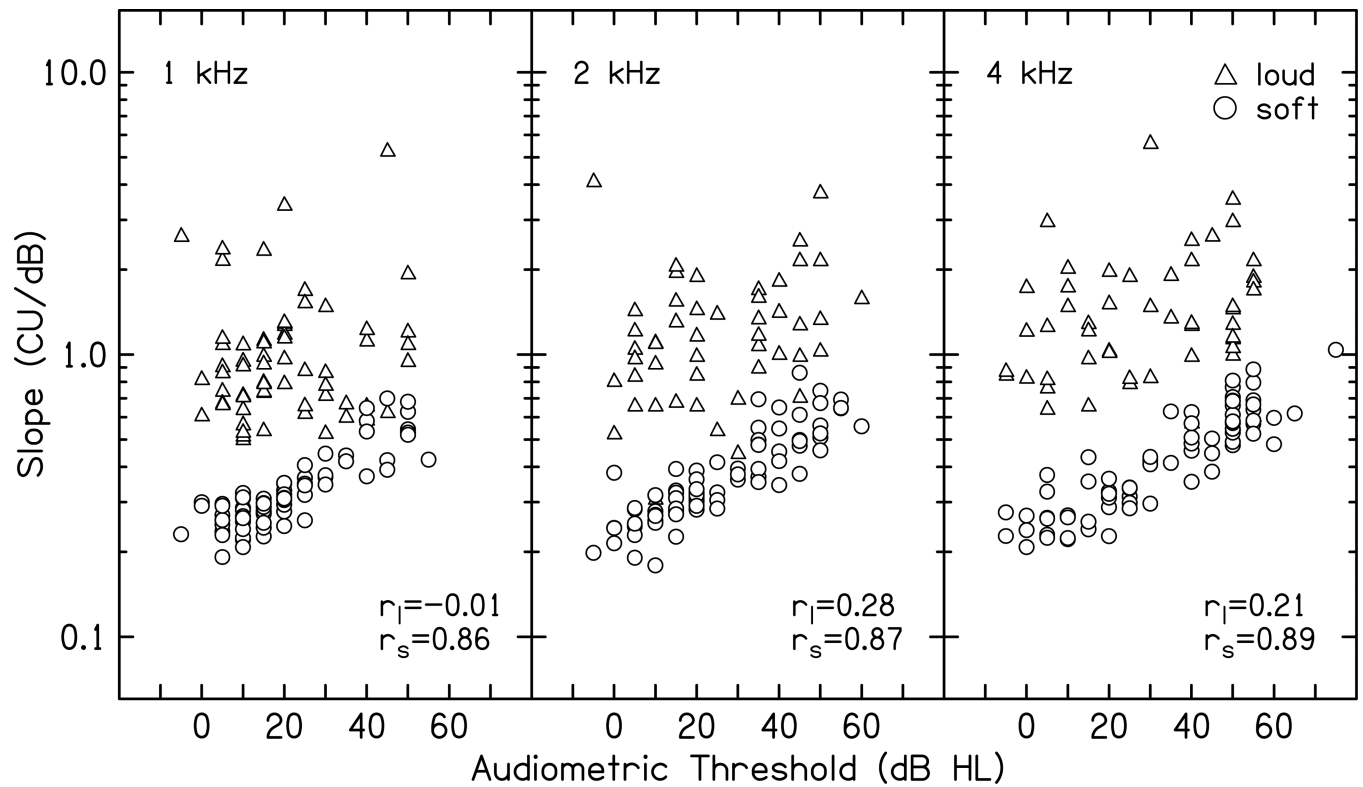


Figure 5. Slopes (CU/dB) of the lines fit to the soft and loud portions of each subject's CLS function as a function of audiometric threshold, with circles representing soft slope and triangles representing loud slope. Data are plotted separately for each frequency.

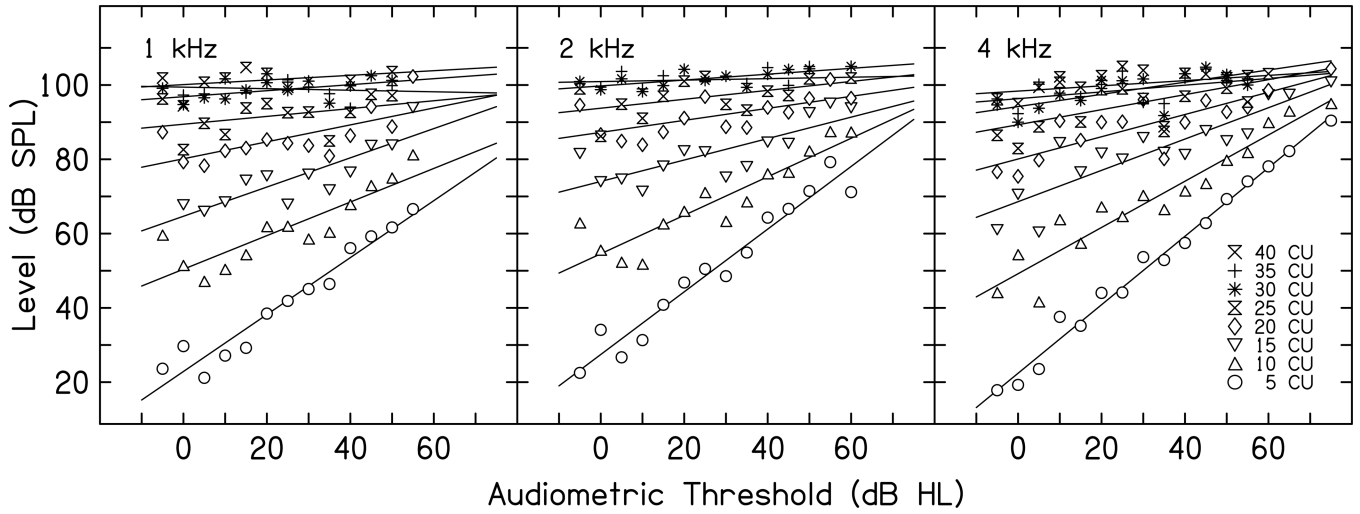


Figure 6. Iso-loudness contours constructed by finding the median level in dB SPL assigned to each loudness category as a function of audiometric-threshold group with CU as the parameter (shown as different symbols). The lines represent best fits to the data for each CU.

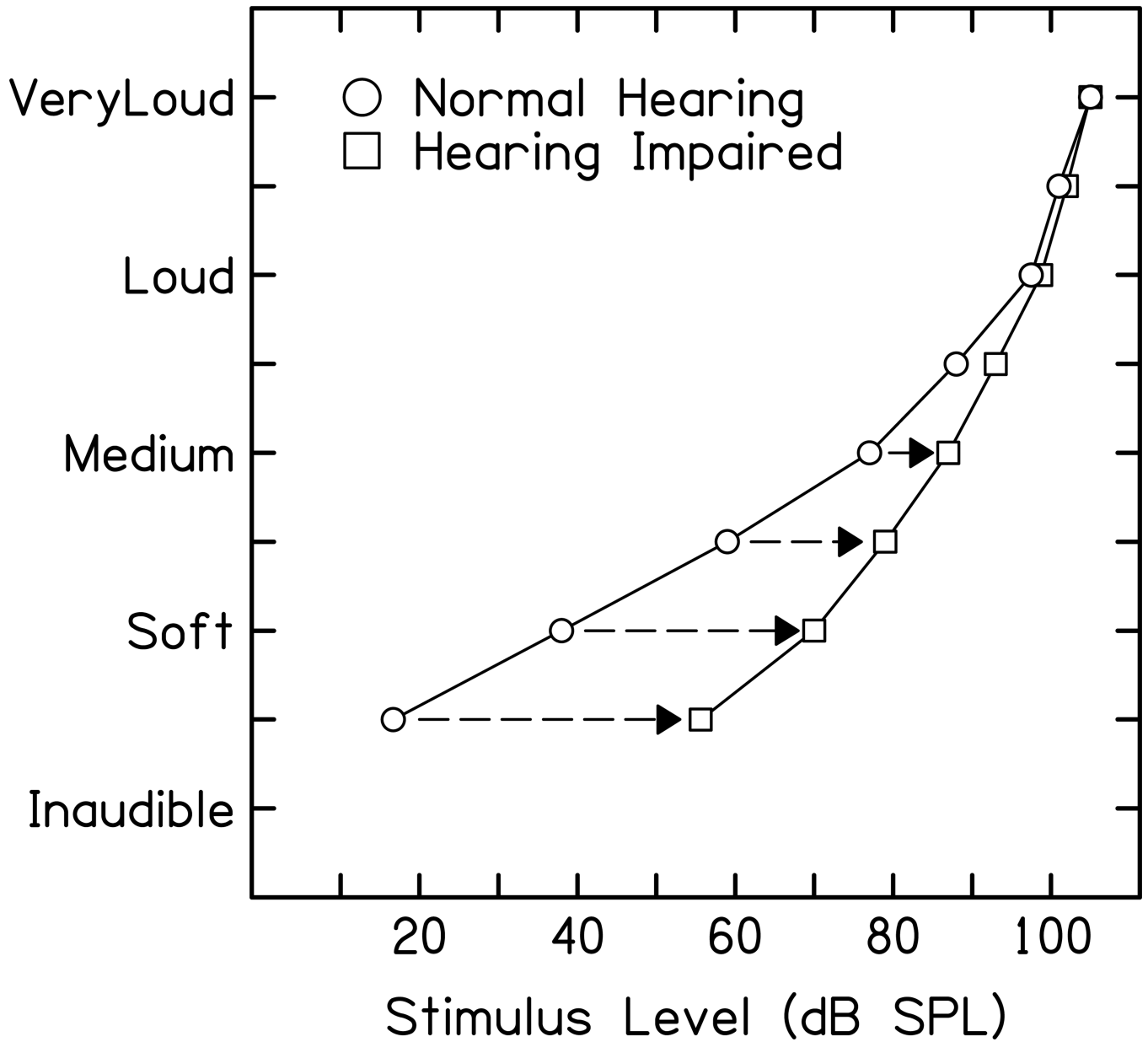


Figure 7. Hypothetical CLS functions for one normal-hearing (circles) and one hearing-impaired subject (squares). The horizontal arrows represent the increase in SPL needed to normalize the loudness percept for the hearing-impaired individual. For ease of visualization, stimulus levels for only half the loudness categories are shown.

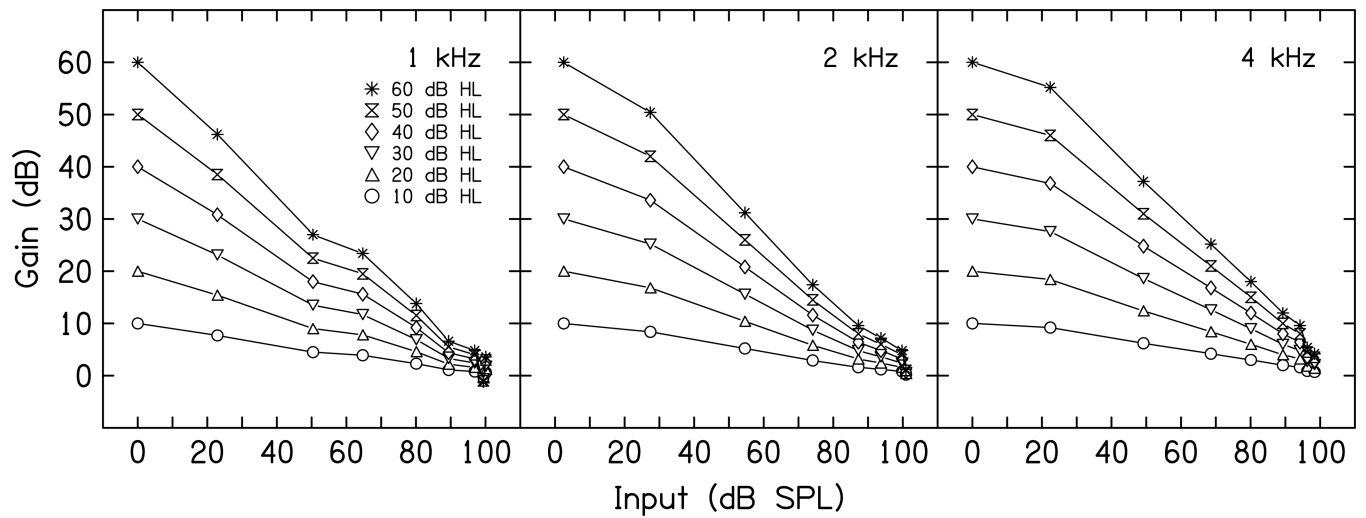


Figure 8.

The estimate of gain needed to normalize the loudness percept as a function of input level with hearing-threshold categories as the parameter. The gain estimates were obtained from the lines fit to the data in Fig. 6. Data for each frequency are shown in separate panels.

Table 1

Mean and median stimulus levels associated with each category were used to determine within-subject reliability across two separate sessions. These comparisons were performed once with outliers included and again with outliers excluded. The reliability of the measurements was assessed using average stimulus-level differences, standard deviations of these differences, and correlations as shown.

	Average Stimulus-level Difference			Standard Deviation of Difference			Correlation Coefficient		
	1 kHz	2 kHz	4 kHz	1 kHz	2 kHz	4 kHz	1 kHz	2 kHz	4 kHz
Mean without outliers	0.16	0.60	-0.61	7.6	6.6	7.8	0.95	0.95	0.94
Mean with outliers	0.17	0.64	-0.65	9.0	6.7	8.2	0.93	0.95	0.93
Median without outliers	0.18	0.81	-0.76	8.2	6.9	8.1	0.94	0.95	0.93
Median with outliers	0.24	0.86	-0.79	9.4	7.1	8.4	0.93	0.95	0.93

Table 2

For each audiometric threshold category, average slopes and average intercepts of the lines fit to the soft portion of each subject's CLS function are provided. Given audiometric threshold (dB HL) and input level (dB SPL), these values can be used to predict an individual's loudness rating when equal to or less than "Medium" (CU=20). Average soft slopes and average intercepts are shown separately for each frequency.

Audiometric Threshold (dB HL)	Average Slope (CU/dB)			Average Intercept (CU)		
	1 kHz	2 kHz	4 kHz	1 kHz	2 kHz	4 kHz
-5	0.23	0.20	0.23	-1.5	-0.5	-1.6
0	0.30	0.27	0.24	-4.4	-3.6	-0.1
5	0.25	0.25	0.28	-0.8	-1.9	-2.7
10	0.27	0.27	0.25	-3.0	-4.7	-4.0
15	0.28	0.30	0.32	-3.9	-8.5	-8.3
20	0.31	0.32	0.31	-7.8	-10.4	-9.2
25	0.34	0.33	0.32	-9.4	-11.9	-9.9
30	0.38	0.38	0.40	-13.1	-13.7	-17.3
35	0.43	0.48	0.52	-15.4	-22.5	-22.7
40	0.53	0.48	0.50	-26.2	-26.5	-25.2
45	0.50	0.55	0.45	-26.2	-32.1	-24.2
50	0.58	0.58	0.62	-33.8	-37.3	-38.4
55	0.42	0.66	0.67	-24.1	-49.0	-43.0
60	--	0.56	0.54	--	-36.2	-37.3
65	--	--	0.62	--	--	-46.3
75	--	--	1.04	--	--	-89.1