A test of the Equal-Loudness-Ratio hypothesis using cross-modality matching functions[a]

Michael Epstein[b]
Institute for Hearing, Speech, and Language and Communications and Digital Signal Processing Center, ECE Department (440 DA), Northeastern University, 360 Huntington Avenue, Boston, Massachusetts 02115

Mary Florentine[c]
Institute for Hearing, Speech, and Language and Department of Speech-Language Pathology and Audiology (133 FR), Northeastern University, 360 Huntington Avenue, Boston, Massachussets 02115

(Received 9 July 2004; revised 13 May 2005; accepted 25 May 2005)

This study tests the Equal-Loudness-Ratio hypothesis [Florentine *et al.*, J. Acoust. Soc. Am. **99**, 1633–1644 (1996)], which states that the loudness ratio between equal-SPL long and short tones is independent of SPL. The amount of temporal integration (i.e., the level difference between equally loud short and long sounds) is maximal at moderate levels. Therefore, the Equal-Loudness-Ratio hypothesis predicts that the loudness function is shallower at moderate levels than at low and high levels. Equal-loudness matches and cross-modality string-length matches were used to assess the form of the loudness function for 5 and 200 ms tones at 1 kHz and the loudness ratio between them. Results from nine normal listeners show that (1) the amount of temporal integration is largest at moderate levels, in agreement with previous studies, and (2) the loudness functions are shallowest at moderate levels. For eight of the nine listeners, the loudness ratio between the 200 and 5 ms tones is approximately constant, except at low levels where it tends to increase. The average data show good agreement between the two methods, but discrepancies are apparent for some individuals. These findings support the Equal-Loudness-Ratio hypothesis, except at low levels. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1954547]

PACS number(s): 43.66.Cb, 43.66.Mk, 43.66.Ba [AJO] Pages: 907–913

I. INTRODUCTION

The Equal-Loudness-Ratio hypothesis states that the loudness ratio between equal-SPL long and short tones is independent of level (Florentine *et al.*, 1996). If the Equal-Loudness-Ratio hypothesis is valid, then loudness functions must be shallower at moderate levels than at low and high levels. This follows because studies have shown that the amount of temporal integration—defined as the level difference between equally loud short and long tones—is greater at moderate levels than at low and high levels (e.g., Florentine *et al.*, 1996, 1998). Support for this notion comes from loudness functions derived from assuming the Equal-Loudness-Ratio hypothesis is true; the loudness functions agree with loudness functions derived from measurements of loudness summation across frequency (Buus, 1999).

To further clarify, if loudness functions for short and long tones are plotted on logarithmic scales with respect to level and are assumed to be parallel, then the horizontal distance between the two functions would indicate the amount of temporal integration at a given level. If a simple power-function model of the growth of loudness is used, temporal integration would have to be constant at all levels because the horizontal distance between the parallel functions would be constant with respect to level. Because measurements of the amount of temporal integration indicate that it varies with level, the slope of the loudness function must also vary in order to create changes in the horizontal distance between the two functions. Specifically, because maximum temporal integration occurs at moderate levels, the slope of the parallel loudness functions must be shallower at those levels.

Some theories, such as the theory of auditory intensity resolution (Durlach and Braida, 1969; Braida and Durlach, 1972) can be applied to attribute the flattening of the loudness function at moderate levels to decisional complexity increases when the test range is large. Ward *et al.* (1996) showed that the slopes of the loudness function decreased when the range was widened using several methods including brightness cross-modality matching. Still, several studies have indicated that loudness functions closely resemble basilar-membrane input/output functions, which are known to be flatter at moderate levels (Buus *et al.*, 1997; Schlauch *et al.*, 1998; Epstein and Florentine, 2005). This makes it unlikely that the mid-level flattening of the loudness functions is only a result of decisional ambiguity resulting from the test range.

The Equal-Loudness-Ratio hypothesis has never been tested directly. Direct measurements of loudness functions for short and long tones are necessary to validate the Equal-Loudness-Ratio hypothesis and its prediction of a shallow

---

[a] A portion of this work was presented at the meeting of the International Society for Psychophysics in 2001 [Florentine, M., Epstein, M., and Buus, S. (2001). "Loudness functions for long and short tones," in Fechner Day 2001, edited by E. Sommerfeld, R. Kompass, and T. Lachmann (Pabst, Berlin).]

[b] Electronic mail: mepstein@ece.neu.edu

[c] Electronic mail: florentin@neu.edu

mid-level segment of the loudness function. Although direct measurements of loudness functions for short tones have been made using magnitude estimation in four listeners (McFadden, 1975), the data vary widely and agree poorly with other investigators' loudness-balance measurements of temporal integration (see Florentine *et al.*, 1996). Given the dearth and uncertainty of magnitude-estimation data for the loudness of short tones, the present study employs a cross-modality matching procedure to measure loudness functions. Such cross-modality matching procedures have been shown to yield reliable data (Teghtsoonian and Teghtsoonian, 1983; Hellman and Meiselman, 1988, 1990, 1993; Hellman, 1999). Additionally, the present study also employs loudness matches between long and short tones to evaluate the transitivity and reliability of the present cross-modality-matching data with another commonly used procedure.

II. METHOD

A. Stimuli

The stimuli were 1 kHz tones with equivalent rectangular durations of 5 and 200 ms. The levels ranged from 5 dB SL to 100 dB SPL for the 200 ms tones and 110 dB SPL for the 5 ms tones. The 1 kHz test frequency and the durations of 5 and 200 ms were chosen to make the measurements directly comparable with several previous studies (Florentine *et al.*, 1996, 1998; Buus *et al.*, 1999). The tones had a 6.67 ms raised-cosine rise and fall. Durations measured between the half-amplitude points were 1.67 ms longer than the nominal durations. Accordingly, the 5 ms stimuli consisted only of the rise and fall, whereas the 200 ms stimuli had a 195 ms steady-state portion. These envelope shapes ensured that almost all the energy of the tone bursts was contained within the 160-Hz-wide critical band centered at 1 kHz (Scharf, 1970; Zwicker and Fastl, 1990). Even for the 5 ms tone burst, the energy within the critical band was only 0.3 dB less than the overall energy.

B. Apparatus

A PC-compatible computer with a signal processor (TDT AP2) generated the stimuli via a 16 bit D/A converter (TDT DD1) with a 50 kHz sample rate. It also recorded the listeners' responses and executed the adaptive procedure. The output of the D/A converter was attenuated (TDT PA4), lowpass filtered (TDT FT5, $\mathrm{fc}=20\ \mathrm{kHz}$, 135 dB/octave), attenuated again (TDT PA4), and led to a headphone amplifier (TDT HB6), which fed one earphone of the Sony MDR-V6 headset. This setup ensured that the stimulus level could be controlled linearly by the attenuators over at least a 130 dB range. For routine calibration, the output of the headphone amplifier was led to an A/D converter (TDT DD1), such that the computer could sample the waveform, calculate its spectrum and rms voltage, and display the results before each block of trials.

C. Procedure

The experiment consisted of three parts. In the first part, absolute thresholds were measured for the test stimuli. In the second part, cross-modality matches were made for the 5 and 200 ms tones to obtain loudness functions. In the third part, equal-loudness matches were made to assess the reliability of the cross-modality matches.

1. Absolute thresholds

Absolute thresholds were measured monaurally for 5 and 200 ms tones at 1 kHz using a two-interval, two-alternative forced-choice paradigm with feedback. On each trial, two observation intervals, marked visually, were presented with an interstimulus interval of 500 ms. The stimulus was presented in the first or second observation interval with equal *a priori* probability for each interval. The listener's task was to press one of two buttons corresponding to the interval containing the stimulus. One hundred milliseconds after the listener's response, the correct answer was indicated by a 200 ms light. Following the feedback, the next trial began after a 500 ms delay.

For each listener and duration, three measurements were made using an adaptive method. A single threshold measurement consisted of three interleaved tracks, each of which ended after five reversals. Reversals occurred when the signal level changed from increasing to decreasing or *vice versa*. On each trial, the track was selected at random among the tracks that had not yet ended. For each track, the level of the signal was initially set approximately 15 dB above the listener's threshold. It decreased following three consecutive correct responses and increased following one incorrect response, such that the signal converged on the level yielding 79.4% correct responses (Levitt, 1971). The step size was 5 dB until the second reversal, after which it decreased to 2 dB.

The threshold for each track was calculated as the average signal level of the fourth and fifth reversals and the average of the three tracks was used as one estimate of the absolute threshold. This method has been shown to provide highly reliable measurements of threshold (Hicks and Buus, 2000). Three such estimates (for a total of nine tracks) were obtained for each listener and duration. The average of the three estimates was used as the reference to set the sensation level, SL, for each listener in the remaining portions of the experiment.

2. Cross-modality matching

Each listener was asked to match the length of a string to the loudness of a sound. The listener was given a virtually unbounded ball of very thin, but strong string (i.e., embroidery floss) and was instructed to cut a piece that was as long as the sound was loud following each stimulus presentation. The 5 and 200 ms tones at the various test levels were presented in mixed order. One block of trials contained five trials at each level and duration. To accustom the listener to the task, a single cross-modality match for each stimulus and level was given as training. Then, two blocks of trials were completed such that ten cross-modality matches were made for each level and duration.

The trials were randomized by selecting each new tone level and duration randomly from a set of possibilities that met the following criteria: the SL needed to be within 30 dB of the level in the previous trial for tones of the same duration and within 25 dB for the other duration. In addition, the stimulus (i.e., a level and duration) must have been presented fewer than five times within the current block of trials. If no stimuli fulfilled these criteria, but some other stimuli still had been presented fewer than five times, a dummy trial was inserted. The dummy trial had the same duration and a level 30 dB above or below the preceding level, depending on the levels of the stimuli that remained to be presented. Such dummy trials were excluded from the data analysis. Large level differences between trials were avoided in order to prevent surprise from a sudden level increase or a missed stimulus from a sudden level decrease.

Each trial was presented in the middle of a 250 ms visually marked interval. After each presentation, the listener cut the string to match the loudness of the presented tone, taped the string segment into a book, turned the page, and pressed a button to indicate completion of the response. The next trial began 700 ms after the listener completed the response. The final cross-modality matches were obtained as the geometric mean of the ten string lengths that were cut to match a given duration and level.

3. Loudness matching

In the final part of the experiment, loudness matches were obtained between 5 and 200 ms tones using a roving-level two-alternative, forced-choice adaptive procedure. This procedure obtained ten concurrent loudness matches by randomly interleaving ten adaptive tracks. Five of these tracks varied the 5 ms tone and five varied the 200 ms tone. The fixed stimulus for each of the five tracks was set to different SLs between 5 and 85 dB in 20 dB steps. (Tracks were eliminated when they exceeded 100 dB SPL for the 200 ms tone or 110 dB SPL for the 5 ms tone.) This procedure ensured that listeners could not identify the stimulus being varied and forced them to use only the two stimuli presented in the current trial to make the loudness judgment (for further discussion, see Buus *et al.*, 1997, 1998).

On each trial, the listener heard two tones separated by a 600 ms interstimulus interval. The fixed-level tone followed the variable tones or the reverse with equal *a priori* probability. The listener's task was to indicate which sound was louder by pressing a key on a response terminal. The next trial began after a 1 s delay. The level of the variable tone was adjusted according to a simple up-down procedure. If the listener indicated that the variable tone was the louder one, its level was reduced, otherwise it was increased. The step size was 5 dB until the second reversal, after which it was 2 dB. This procedure made the variable tone converge toward a level at which it was judged louder than the fixed tone in 50% of the trials (Levitt, 1971).

For each track, the variable stimulus was initially set approximately 15 dB below the expected equal-loudness level. (If this was below threshold, the variable stimulus was set to threshold.) This starting level ensured that the listener would initially hear some trials in which the short tone was definitely louder and some trials in which the long tone was definitely louder. On each trial, the track was chosen at random among those that had not yet ended, which they did after nine reversals. The average level of the last four reversals of each track was used as an estimate of the level at which the loudness of the variable tone was equal to that of the fixed-level tone. Three such loudness matches were obtained for each fixed stimulus and the average was used to estimate the level difference needed for equal loudness. The level difference between short and long tones that are judged to be equally loud will be compared to the level difference between short and long tones that yield equal string lengths.

D. Listeners

A total of nine listeners were tested on all conditions. All listeners had bilaterally normal thresholds and medical histories consistent with normal hearing. They ranged in age from 20 to 46 years. Test ears had audiometric thresholds within 10 dB of ANSI (1989) standard at octave frequencies from 250 to 8000 Hz. Most of the listeners had previous experience making loudness judgments, except for L5 and L7. Listener L4 is the first author.

E. Data analysis

For each listener, one data point was the geometric mean of ten string lengths. The standard error of the mean was determined from the logarithms of the string lengths. The group mean and standard deviations were calculated across the nine individual listeners' geometric means for each duration and SL. The resulting data were transformed back into the string-length domain to show the average and probable range of individual listeners' responses.

To examine the effects of stimulus variables, an analysis of variance (ANOVA) for repeated-measures was performed on the logarithms of the ten string lengths obtained for each listener, SL, and duration. Because the Equal-Loudness-Ratio hypothesis states that the loudness ratio between equal-SPL long and short tones is independent of SPL, the SLs were transformed into approximate "SPLs" for each listener to obtain an indicator of the stimulus level for the ANOVA. However, loudness functions are more similar across listeners when evaluated in terms of SL than in terms of SPL (Hellman and Zwislocki, 1961). In other words, using true SPLs would obscure similarities across listeners. To keep the loudness functions for the different listeners aligned according to SL, while ensuring that the 5 and 200 ms loudness functions were aligned according to approximately equal SPLs, 0 dB SL for the 200 ms tones was equated with 10 dB "SPL" for all listeners. (The true average of all the listeners' thresholds was 9.9 dB SPL.) For the 5 ms tones, each listener's threshold difference between the 5 and 200 ms tones was rounded to the nearest 5 dB and added to 10 dB "SPL" to obtain the "SPL" corresponding to 0 dB SL. For eight of the nine listeners, the rounded threshold difference of 15 dB was between 2 dB higher and 1 dB lower than the true differences. For the ninth listener, the threshold difference was exactly 10 dB. To ensure that the ANOVA encompassed only "SPLs" for which data were available for both 5 and 200 ms

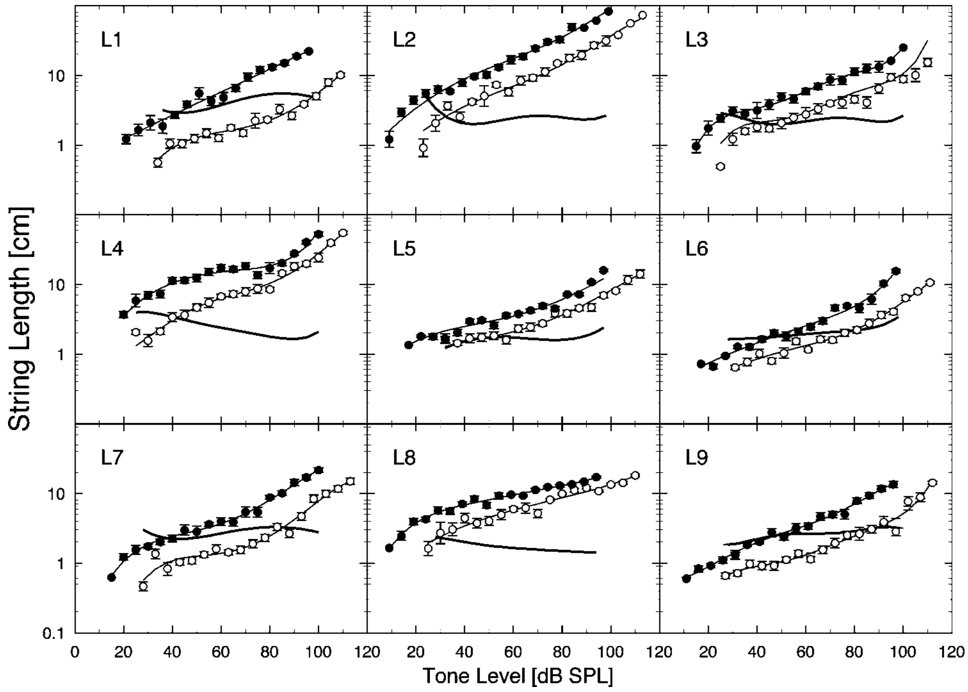

FIG. 1. Individual CMM functions obtained from the nine listeners. The geometric mean of string length is plotted on a log scale as a function of level. The closed circles show data for the 200 ms tones and the open circles show the data for the 5 ms tones. The vertical bars show ± one standard error of the mean. The thin lines show fourth-order polynomials fitted to the data. The thick lines show the ratio of string lengths obtained for equal-SPL 200 and 5 ms tones as estimated from the polynomials.

tones, the analyses encompassed only "SPLs" between 30 and 95 dB. In the following, the sound level corrected in this manner will be referenced simply as SPL.

An initial ANOVA showed that repetition and all interactions with it were not significant factors. Therefore, the main analyses did not include repetition as a factor. The primary ANOVA examined the effects of level and duration. In this analysis, listener was used as a random factor to produce a repeated-measures analysis of variance. For all tests, the outcome was considered significant when $p \leq 0.05$.

III. RESULTS

Figure 1 shows the individual cross-modality-matching (CMM) functions obtained from the nine listeners. The geometric mean of string length is plotted on a log scale as a function of level. The full range of string lengths cut was from 0.1 to 159.1 cm. The data for individual listeners are generally consistent, as indicated by the small standard errors and the general monotonicity. However, there are clear differences among listeners as is typical of measurements of loudness. Despite the individual differences, a few findings are clear. For all levels, the string length matched to the short tone is less than that for the long tone in agreement with classic temporal-integration data (for review, see Florentine *et al.*, 1996). For most of the listeners, the cross-modality-matching functions for both the 5 and 200 ms tones are shallower at moderate levels than at low and high levels. They are also nearly parallel, as indicated by the roughly constant vertical distance between the two functions. The thick, solid lines in Fig. 1, show the ratio of string lengths matched to equal-SPL long and short tones. It is approximately independent of SPL for most of the listeners, although some exceptions are apparent.

The average CMM functions are shown in Fig. 2. They are plotted in the same manner as Fig. 1. The average data show the same general trends as the majority of the individual data. The ratio of string lengths matched to equal-SPL long and short tones is approximately independent of SPL, except for a slight increase below 40 dB SPL. Like the individual data, both loudness functions are shallower at moderate levels than at low and high levels.

These observations are supported by the ANOVA. The effects of SPL and duration are both highly significant ($P < 0.0001$), but the interaction between them is not ($P = 0.68$), as is expected if the effect of duration is independent of SPL. Accordingly, the ANOVA is consistent with the Equal-Loudness-Ratio hypothesis.

Although individual and group data for the CMM functions appear quite orderly, the comparison with the direct loudness matches used to validate them shows considerable variability among the listeners. Figure 3 compares individual listeners' adaptive loudness matches with indirect loudness matches obtained from their CMM data. It shows the amount

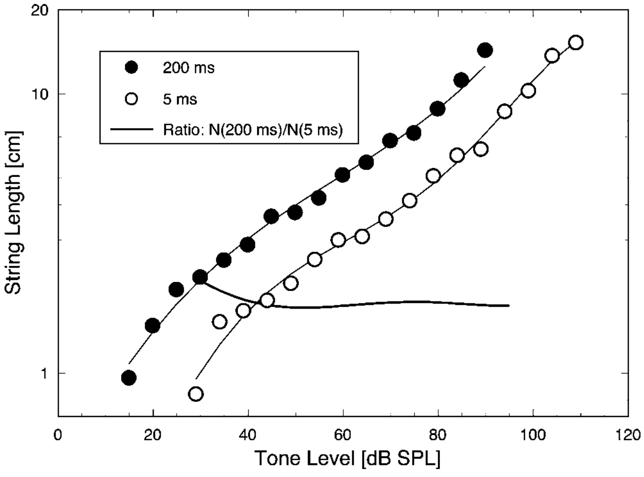

FIG. 2. Mean data and fourth-order polynomials fitted to the data plotted in Fig. 1. The thick line shows the ratio of string lengths obtained for equal-SPL 200 and 5 ms tones as averaged from the polynomials.

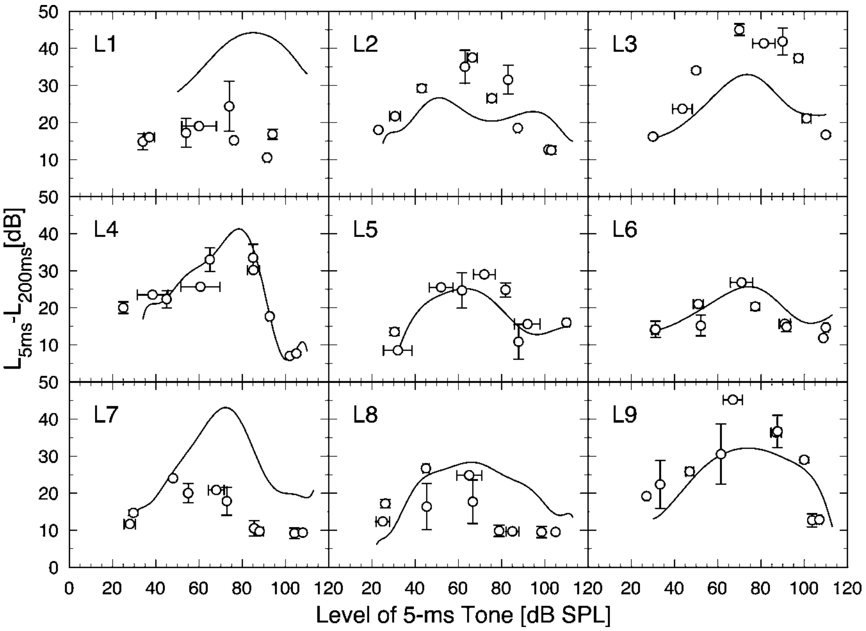

FIG. 3. Temporal integration of loudness for nine individual listeners derived from adaptive loudness (circles) and CMM (lines). The level differences needed to obtain equal loudness between the 5 and 200 ms tones are plotted as a function of the level of the 5 ms tone. The error bars show ± one standard error of the mean.

of temporal integration—defined as the level difference between equally loud 5 and 200 ms tones—plotted as a function of the 5 ms tone's level. Each panel shows data for one listener. The circles show the amount of temporal integration obtained with the adaptive loudness-matching procedure. The error bars indicate the standard error of the mean and are oriented to indicate which tone level was varied. The solid line shows the amount of temporal integration derived from the individual CMM data. The latter function was obtained as the level differences between 5 and 200 ms tones that yielded equal string lengths according to polynomials fitted to the logarithms of the geometric means for each listener and duration. All listeners show a mid-level maximum in the amount of temporal integration in agreement with a number of previous studies (e.g., Florentine *et al.*, 1996, 1998; Buus, 1999). The mid-level maximum is clearly present in the individual data, both for adaptive loudness matches and for the indirect matches obtained from the cross-modality matches. The magnitude of the mid-level maximum varies greatly among individuals. Of the nine listeners, four [L4, L5, L6, and L9 (except for one point)] show excellent agreement between the cross-modality matches and the equal-loudness matches. Four listeners (L2, L3, L7, and L8) show agreement at some levels and one listener (L1) shows quite large disagreement. For the four listeners who show agreement at some levels there are some data points that clearly do not overlap between the adaptive matching procedure and the CMM procedure.

Despite substantial differences between the two sets of data for some listeners, average data were calculated to compare with other average data in the literature. Figure 4 shows the average amount of temporal integration across the nine listeners plotted in the same manner as Fig. 3. The standard deviation across the individual listeners is very large, in contrast to the variability for the individual listeners. Although there are some differences between the two methods at moderate and high levels, the average results show reasonable agreement. This indicates that the CMM and loudness-matching procedures produce a generally consistent assessment of the relation between the loudness of short and long tones for group data, except at the highest levels.

IV. DISCUSSION

A. Agreement between cross-modality matching and loudness matches

Agreement between the measurements of CMM and the adaptive loudness balances varies from reasonable to excellent for seven of the nine listeners. Although it is difficult to estimate the probable error of the amount of temporal integration derived from the CMM functions, the data for two other listeners (L1 and L7) appear to show substantial discrepancies between the two methods. One possible reason for these discrepancies is that estimates of the amount of temporal integration from the CMM data are quite sensitive to perturbations in the polynomial fits used to summarize the CMM data. This is especially true when the slope of the CMM function is shallow, as it is at moderate levels for most

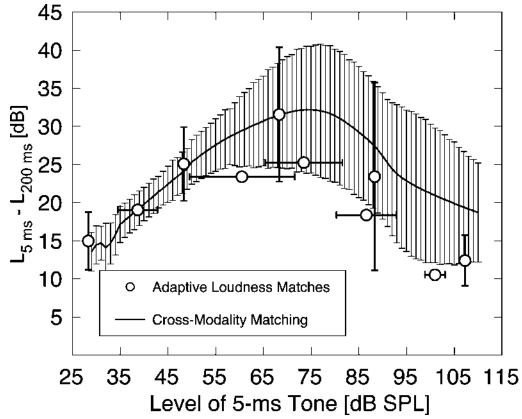

FIG. 4. Average temporal integration of loudness for nine listeners plotted in the same manner as Fig. 3. The error bars show ± one standard deviation calculated across the nine listeners' data for the loudness matching experiment.

listeners, including L7 and L1 (primarily for 5 ms tones). Assuming that variation is uniform across the entire intensity range, a shallower slope increases the sensitivity because a wider range of stimulus levels will result in a statistically identical cross-modality match.

The idea that random error is responsible for most of the discrepancies between CMM and loudness matches noted for individual listeners is supported by the reasonable agreement between the two data sets shown by the average data in Fig. 4 except at high levels. Whereas the two sets of data agree up to about 55 dB SPL, differences are apparent at higher levels. These differences may be due to effects of induced loudness reduction [ILR; Nieder *et al.*, 2003; also called loudness recalibration (Marks, 1994), as described in the following]. Therefore, the difference between the two data sets at high levels should not be taken to indicate that one or the other method yields invalid results, at least when data are averaged across listeners.

Generally, ILR refers to the finding that intense tones reduce the loudness of subsequent weaker tones at or near the same frequency (e.g., Marks, 1994; Nieder *et al.*, 2003; Arieh and Marks, 2003). Recently, it has been shown that ILR is likely to affect measurements of temporal integration of loudness considerably when the stimulus level varies from trial to trial, as it did for both methods used to assess loudness in the present study. Nieder *et al.* (2003) found that 5 ms tones are much less effective in inducing ILR on 200 ms tones than the reverse. In addition, ILR effects are greatest when inducers are at moderate-to-high levels and relatively close in level to the test tone. (The maximum effect occurs with approximately a 10 dB separation.) Thus, when 5 and 200 ms tones at various levels separated by relatively small level differences are presented within a block of trials, the loudness reduction is probably greater for 5 ms tones than for 200 ms tones. In turn, the SPL of the 5 ms tone must be raised more to achieve the same loudness as a 200 ms tone, which leads to an enlarged estimate of the amount of temporal integration.

Because it is highly likely that a high-level presentation of a long tone would have occurred early in the experimental block, the data of Nieder *et al.* (2003) indicate that ILR would have affected both the loudness matches and the cross-modality matches. However, the CMM procedure contained more levels with closer spacing than the loudness-matching procedure. Furthermore, because the present CMM procedure restricted the amount of level change from trial to trial, the listener received moderate and high stimuli closer together in time, offering less opportunity for recovery than in the loudness matches. Therefore, the effect of ILR could be greater for CMM than for adaptive loudness matching. Individual differences in performance on psychophysical tasks make it difficult to see a clear effect of procedure in the individual data. Further understanding of the time course of ILR would be necessary to determine what the precise effect might have been.

B. Comparison with data in the literature

To facilitate the comparison with previous studies, it is useful to approximate the mid-to-high-level portion of the CMM functions with power functions. Above 40 dB SPL, the data for both 5 and 200 ms tones approximate a power function with an exponent of about 0.14 and the ratio between them is approximately 1.8.

The exponent of 0.14 is somewhat smaller than the 0.2 CMM exponent reported by Baird *et al.* (1980) and considerably smaller than the 0.32 reported for matches between line length and loudness (Hellman, 1999). However, some discrepancy is to be expected because Hellman's (1999) listeners adjusted tones to match fixed-length lines, whereas the present listeners had to match a variable string length to a given tone. Given the general tendency of judgments to regress toward the middle of the scale, one would expect the present cross-modality functions to have lower exponents than those obtained by Hellman (1999). Likewise, the ratio 1.8 is considerably smaller than 4.0 estimated in recent loudness-balance experiments (e.g., Florentine *et al.*, 1998; Buus *et al.*, 1999). However, if the present ratio is scaled by the ratio between the present exponent and that of about 0.3 generally used to approximate the growth of loudness at moderate and high levels, the corresponding loudness ratio is about 3.9, which is in excellent agreement with the ratios estimated in loudness-balance experiments.

C. Testing the Equal-Loudness-Ratio hypothesis

The present group data show that the ratio of string lengths matched to equal-SPL long and short tones is approximately constant, except for a slight increase below 40 dB SPL. Accordingly, the data support the Equal-Loudness-Ratio hypothesis, except at low levels.

The finding that the loudness ratio between equal-SPL long and short tones is approximately constant is not unexpected. This relationship is an inherent property of Zwislocki's (1969) theory of temporal integration. Moreover, assuming that loudness bears a simple relation to the overall neural activity evoked by the stimulus, this finding agrees with data on auditory-nerve adaptation. Smith and Zwislocki (1975) showed that the ratios of spike rates measured in the auditory nerve at various times after the onset of a stimulus were approximately independent of the spike rate. This finding indicates that the ratio between the number of spikes evoked by equal-SPL long and short tones is approximately independent of their SPL. Thus, one would expect that the loudness ratio is also independent of SPL, even if loudness may be formed after central transformations of the auditory-nerve activity and may not be directly proportional to nerve-spike count (Relkin and Doucet, 1997). As discussed by Buus and Florentine (2001), the loudness ratio appears to be nearly proportional to the integral of the square of the firing rate in the auditory nerve, rather than being equal to the ratio between the numbers of spikes evoked by the stimuli (Zeng and Shannon, 1994). Nevertheless, it is clear that the present data regarding how loudness changes with duration agree with expectations based on auditory-nerve data and both the present data and auditory-nerve data support the Equal-Loudness-Ratio hypothesis.

V. CONCLUSIONS

The CMM procedure used to measure loudness functions for long and short tones yields reasonably reliable results for most listeners. Group comparisons with loudness-matching data for the same listeners and stimuli indicate that the listeners' average loudness judgments were generally internally consistent, although the individual data are variable. The loudness functions obtained by CMM generally supported the Equal-Loudness-Ratio hypothesis. The loudness functions also show a decrease in slope at moderate levels consistent with previous studies.

ACKNOWLEDGMENTS

Søren Buus contributed substantially to this project and passed away prior to completion of the work. Rhona Hellman, Bert Scharf, and Eva Wagner gave helpful comments on an earlier version of this manuscript. Editor Andrew Oxenham and reviewers Bert Schlauch and Lawrence Ward also provided helpful suggestions. This research was supported by NIH/NIDCD Grant No. R01DC02241.

ANSI (1989). "Specifications for audiometers," ANSI S3.6-1989.

Arieh, Y. and Marks, L. E. (**2003**). "Time course of loudness recalibration: Implications for loudness enhancement," J. Acoust. Soc. Am. **114**, 1550–1556.

Baird, J. C., Green, D. M., and Luce, R. D. (**1980**). "Variability and sequential effects in cross-modality matching of area and loudness," J. Exp. Psychol. Hum. Percept. Perform. **6**, 277–289.

Braida, L. D. and Durlach, N. I. (**1972**). "Intensity perception. II. Resolution in one-interval paradigms," J. Acoust. Soc. Am. **51**, 483–502.

Buus, S. (**1999**). "Loudness functions derived from measurements of temporal and spectral integration of loudness," in *Auditory Models and Non-linear Hearing Instruments*, edited by A. N. Rasmussen, P. A. Osterhammel, T. Andersen, and T. Poulsen (GN ReSound, Taastrup, Denmark).

Buus, S. and Florentine, M. (**2001**). "Modifications to the power function for loudness," in *Fechner Day 2001*, edited by E. Sommerfeld, R. Kompass, and T. Lachmann (Pabst, Berlin).

Buus, S., Florentine, M., and Poulsen, T. (**1997**). "Temporal integration of loudness, loudness discrimination, and the form of the loudness function," J. Acoust. Soc. Am. **101**, 669–680.

Buus, S., Florentine, M., and Poulsen, T. (**1999**). "Temporal integration of loudness in listeners with hearing losses of primarily cochlear origin," J. Acoust. Soc. Am. **105**, 3464–3480.

Buus, S., Müsch, H., and Florentine, M. (**1998**). "On loudness at threshold," J. Acoust. Soc. Am. **104**, 399–410.

Durlach, N. I. and Braida, L. D. (**1969**). "Intensity perception. I. Preliminary theory of intensity resolution," J. Acoust. Soc. Am. **46**, 372–383.

Epstein, M. and Florentine, M. (**2005**). "Inferring basilar-membrane motion from tone-burst otoacoustic emissions and psychoacoustic measurements," J. Acoust. Soc. Am. **117**, 263–274.

Florentine, M., Buus, S., and Poulsen, T. (**1996**). "Temporal integration of loudness as a function of level," J. Acoust. Soc. Am. **99**, 1633–1644.

Florentine, M., Buus, S., and Robinson, M. (**1998**). "Temporal integration of loudness under partial masking," J. Acoust. Soc. Am. **104**, 999–1007.

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. (**1988**). "Prediction of individual loudness exponents from cross-modality matching," J. Speech Hear. Res. **31**, 605–615.

Hellman, R. P. and Meiselman, C. H. (**1990**). "Loudness relations for individuals and groups in normal and impaired hearing," J. Acoust. Soc. Am. **88**, 2596–2606.

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.

Hellman, R. P. and Zwislocki, J. J. (**1961**). "Some factors affecting the estimation of loudness," J. Acoust. Soc. Am. **33**, 687–694.

Hicks, M. L. and Buus, S. (**2000**). "Efficient across-frequency integration: Evidence from psychometric functions," J. Acoust. Soc. Am. **107**, 3333–3342.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.

Marks, L. E. (**1994**). " 'Recalibrating' the auditory system: The perception of loudness," J. Exp. Psychol. Hum. Percept. Perform. **20**, 382–396.

McFadden, D. (**1975**). "Duration-intensity reciprocity for equal loudness," J. Acoust. Soc. Am. **57**, 702–704.

Nieder, B., Buus, S., Florentine, M., and Scharf, B. (**2003**). "Interactions between test- and inducer-tone durations in induced loudness reduction," J. Acoust. Soc. Am. **114**, 2846–2855.

Relkin, E. M. and Doucet, J. R. (**1997**). "Is loudness simply proportional to the auditory nerve spike count?," J. Acoust. Soc. Am. **101**, 2735–2740.

Scharf, B. (**1970**). "Critical bands," in *Foundations of Modern Auditory Theory*, edited by J. V. Tobias (Academic, New York), Vol. I.

Schlauch, R. S., DiGiovanni, J. J., and Ries, D. T. (**1998**). "Basilar membrane nonlinearity and loudness," J. Acoust. Soc. Am. **103**, 2010–2020.

Smith, R. L. and Zwislocki, J. J. (**1975**). "Short-term adaptation and incremental responses in single auditory-nerve fibers," Biol. Cybern. **17**, 169–182.

Teghtsoonian, M. and Teghtsoonian, R. (**1983**). "Consistency of individual exponents in cross-modal matching," Percept. Psychophys. **33**, 203–214.

Ward, L. M., Armstrong, J., and Golestani, N. (**1996**). "Intensity resolution and subjective magnitude in psychophysical scaling," Percept. Psychophys. **58**, 793–801.

Zeng, F.-G. and Shannon, R. V. (**1994**). "Loudness-coding mechanisms inferred from electric stimulation of the human auditory system," Science **264**, 564–566.

Zwicker, E. and Fastl, H. (**1990**). *Psychoacoustics - Facts and Models* (Springer, Berlin).

Zwislocki, J. J. (**1969**). "Temporal summation of loudness: An analysis," J. Acoust. Soc. Am. **46**, 431–441.