LOUDNESS FUNCTIONS FOR LONG AND SHORT TONES

Mary Florentine,[1,2] Michael Epstein,[1,3] and Søren Buus[1,3]
[1]Institute for Hearing, Speech, and Language
[2]Dept. of Speech-Language Pathology and Audiology (133 FR)
[3]Communications and Digital Signal Processing Center, ECE Dept. (440 DA)
Northeastern University, 360 Huntington Avenue, Boston, MA 02115 U.S.A.
E-mail: florentin@neu.edu, mepstein@ece.neu.edu, buus@neu.edu

Abstract

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.*

Knowledge about auditory processing of brief sounds is essential for understanding the perception of speech and environmental sounds. Despite the dominance of short sounds in the natural environment, most psychoacoustic experiments employ long-duration stimuli. The purpose of the present study is to gain a better understanding of the loudness of brief sounds and how it relates to that of long sounds.

The Equal-Loudness-Ratio hypothesis suggests that a simple, direct relationship exists between the loudness of long and short sounds. It states that the loudness ratio between equal-SPL long and short tones is independent of level (Florentine *et al.*, 1996). This relationship is an inherent property of Zwislocki's (1969) theory of temporal integration and agrees with data on auditory-nerve adaptation (Smith and Zwislocki, 1975). 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 a number of 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;

Florentine *et al.*, 1998; Buus *et al.*, 1997; Buus *et al.*, 1999; Buus, 1999). However, the Equal-Loudness-Ratio hypothesis has never been tested directly.

Measurements of loudness functions are necessary to validate the Equal-Loudness-Ratio hypothesis and its prediction of a shallow mid-level segment of the loudness function. Therefore, a cross-modality matching procedure was used as a direct measure of loudness. Additionally, loudness matches between long and short tones were made to demonstrate the transitivity and reliability of the cross-modality-matching procedure.

Method

For all three parts of the experiment, nine listeners with normal hearing were presented 1-kHz tones via earphones. Measurements were made monaurally in a double-walled sound-attenuating booth.

1. Absolute Thresholds
Absolute thresholds were measured for 5- and 200-ms tones using a transformed up-down method and 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 either the first or second observation interval with equal *a priori* probability for each interval. The listener's task was to indicate the interval containing the stimulus by pressing one of two buttons. 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.

A single threshold measurement consisted of three interleaved adaptive 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 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 considered a single absolute-threshold measurement. Three such measurements (for a total of nine tracks) were obtained for each listener and duration. The average across all measurements was used as the reference for setting the sensation level for each listener in the remaining portions of the experiment.

2. Cross-Modality Matching
Cross-modality matching was performed by asking the listener to match the length of a string to the loudness of a sound. The listener was given a virtually unbounded ball of embroidery floss and was instructed to "cut a piece that was as long as the sound is loud following each stimulus presentation." No reference or range was given as a basis for this judgment. The 5- and 200-ms tones were presented in mixed order at levels encompassing the range from 5 dB SL to 110 dB SPL for the 5-ms tones and 5 dB SL to 100 dB SPL for the 200-ms tones in 5-dB steps. One block of trials contained five trials at each level and duration. To accustom the listener to the task, a single match for each level and duration was performed as training.

Then, two blocks of trials were completed such that ten matches were obtained for each level and duration.

The trials were chosen by selecting each new tone level and duration randomly from the 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 level and duration pair 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. The results of dummy trials were not included in the final analysis.

Each stimulus was presented in the middle of a 250-ms interval, which was marked visually. After each presentation, the listener cut a piece of string to match the loudness and pressed a button to indicate completion of the response. After the listener completed the response, the next trial began after a 700-ms delay. The final measurement was calculated as the geometric mean of the ten matches (string lengths) completed for each stimulus.

3. Loudness Matching

The final part of the experiment consisted of loudness matches between 5- and 200-ms tones. This was performed using a roving-level two-alternative forced-choice adaptive procedure (Buus *et al.*, 1999). This procedure obtains ten concurrent loudness matches by using ten randomly interleaved 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. (If 85 dB SL exceeded 100 dB SPL for the 200-ms tone or 110 dB SPL for the 5-ms tone, that track was omitted.) 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 a loudness judgment.

On each trial, the listener heard two tones separated by a 600-ms interstimulus interval. The fixed-level tone followed the variable tone 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 and 2 dB thereafter. This procedure made the variable tone converge towards 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 louder and some trials in which the long tone was 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. The final estimate of each stimulus pair was taken as the average of three such loudness matches. (The amounts of temporal integration obtained in this manner will be compared to those derived from the cross-modality-matching experiment.)

Results and Discussion

Figure 1 shows the average loudness functions obtained from cross-modality matches by nine listeners. The geometric mean of string length is plotted on a log scale as a function of level. For all levels, the loudness of the short tone is less than the loudness of the long tone in agreement with classic temporal-integration data (for review, see Florentine *et al.*, 1996). Furthermore, the loudness functions for 5- and 200-ms tones are nearly parallel and both are shallower at moderate levels than at low and high levels. The vertical distance between the two functions indicates the ratio of string lengths matched to equal-SPL long and short tones. The dashed line in Fig. 1 shows this ratio. It is approximately constant except for a slight increase below 40 dB SPL. Accordingly, the present data support the Equal-Loudness-Ratio hypothesis except at low levels.

Above 40 dB SPL, both cross-modality-matching functions 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 considerably smaller than that 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 our 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 that obtained by Hellman (1999). Likewise, the ratio

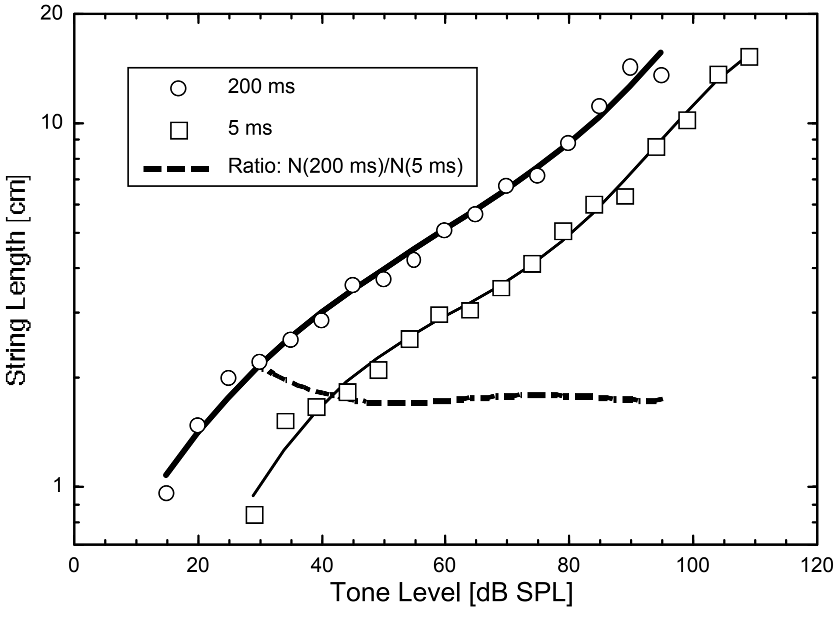

Figure 1. Geometric mean of string lengths that nine listeners matched to the loudnesses of long and short tones in the cross-modality-matching experiment. The solid lines show fourth-order polynomials fitted to the string-length data. The dashed line shows the ratio of string lengths obtained for equal-SPL 200- and 5-ms tones as estimated from the polynomials.

of 1.8 is considerably smaller than that of about 4 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.

To evaluate the consistency of the listeners' loudness judgments, cross-modality matches were compared with adaptive loudness matches. Figure 2 shows the average level difference between equally loud 5- and 200-ms tones plotted as a function of the 5-ms tone's level. The unfilled circles show the amounts of temporal integration obtained with the adaptive loudness-matching procedure. The solid line shows the amount of temporal integration derived from the individual cross-modality-matching data. The latter function was obtained by averaging 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. Although there are some differences between the two methods at moderate and high levels, the results show reasonable transitivity, which indicates that the cross-modality-matching procedure produces valid and reliable results.

In conclusion, the mid-level flattening of the cross-modality matching functions shown in Fig. 1 and the mid-level maximum of the amounts of temporal integration shown in Fig. 2 are likely to characterize these listeners' perception of loudness. Accordingly, the loudness functions agree with the Equal-Loudness-Ratio hypothesis, except at low levels where the

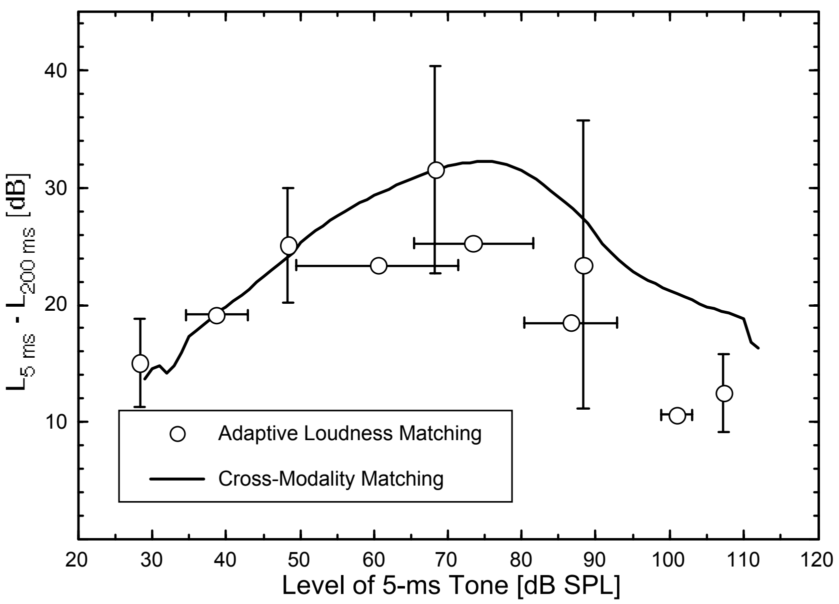

Figure 2. Average amounts of temporal integration of loudness for nine listeners derived from adaptive loudness matching and cross-modality matching. The error bars indicate the standard deviation calculated across the nine listeners' data for the loudness-matching experiment.

loudness ratio between equal-SPL long and short tones appears to increase somewhat.
The finding that the loudness ratio between equal-SPL long and short tones is approximately constant is not unexpected. It is inherent in Zwislocki's (1969) theory of temporal summation of loudness. Moreover, assuming that loudness bears a simple relation to the overall neural activity evoked by the stimulus, it 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 also is independent of SPL, even if loudness may be formed after central transformations of the auditory-nerve activity. As discussed by Buus and Florentine (this volume), the loudness ratio appears to be equal 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. Nevertheless, it is clear that the present data agree with expectations based on auditory-nerve data and both support the Equal-Loudness-Ratio hypothesis.

Summary

Loudness functions for long and short tones were measured using a cross-modality-matching procedure. This procedure yielded reasonably reliable results. Comparisons with loudness-matching data for the same listeners and stimuli indicate that the listeners' average loudness judgments were internally consistent. The loudness functions obtained by cross-modality matching generally supported the Equal-Loudness-Ratio hypothesis. They showed that the loudness ratio between equal-SPL long and short tones is independent of SPL except at low levels. The loudness functions also showed a decrease in slope at moderate levels. The shallower mid-level slopes were consistent with the expected form of the loudness functions generated from the Equal-Loudness-Ratio hypothesis given the mid-level maximum of the amount of temporal integration shown by previous studies.

Acknowledgments

This research was supported by NIH/NIDCD Grant No. R01DC02241.

References

Buus, S. (1999). Loudness functions derived from measurements of temporal and spectral integration of loudness. In: A. N. Rasmussen, P. A. Osterhammel, T. Andersen and T. Poulsen (Eds.), Auditory models and non-linear hearing instruments, 135-188. Taastrup, Denmark: GN ReSound.

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.

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 Hearing*, **20**, 193-213.

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

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

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