

## MEASURING LOUDNESS OF LONG AND SHORT TONES USING MAGNITUDE ESTIMATION

Michael Epstein<sup>1,2</sup>, Mary Florentine<sup>1,3</sup>, and Søren Buus<sup>1,2</sup>

<sup>1</sup>Institute for Hearing, Speech, and Language

<sup>2</sup>Communications and Digital Signal Processing Center, ECE Dept. (440 DA)

<sup>3</sup>Dept. of Speech-Language Pathology and Audiology (133 FR)

Northeastern University, 360 Huntington Ave., Boston, MA 02115 U.S.A.

Email: mepstein@ece.neu.edu, florentin@neu.edu, buus@neu.edu

### Abstract

*McFadden (1975) questioned the accuracy and reliability of the method of magnitude estimation for the measurement of loudness of tones that vary both in duration and level. He suggested that it produced unreliable results and should not be used to assess how loudness depends on stimulus duration. To examine this issue further, the present study obtained loudness functions for 5- and 200-ms tones in nine listeners using magnitude estimation. Loudness matches between the short and long tones were also obtained using an adaptive 2IFC procedure. The average amounts of temporal integration (defined as the level difference between equally loud short and long tones) obtained with the two procedures show good agreement. However, this may not apply to individual listeners. Some listeners show poor agreement, whereas others show good agreement. These results indicate that magnitude estimation provides a rapid and accurate means for assessing group-average loudness functions for tones of different durations. Nonetheless, it appears that magnitude estimation often is too susceptible to judgment biases and variability to reveal detailed features of individual listeners' loudness functions for tones of various durations, even if it does reveal their general form.*

Magnitude estimation has frequently been used to measure the growth of loudness. As a procedure, it has received its share of criticism (e.g., McFadden, 1975; Poulton, 1989) as well as praise (e.g., Stevens, 1975; Hellman and Zwislocki, 1964).

Previous experiments using magnitude estimation to assess temporal integration have had mixed success. Stevens and Hall (1966) attempted to obtain information about the growth of loudness for bursts of white noise with different durations using magnitude estimation. They measured loudness across a relatively wide range of levels (36 to 109 dB SPL) and assumed a simple power-function representation. They found that the average slope of the loudness functions were nearly the same for noise bursts ranging in duration from 5 to 500 ms. They did not report on the variability, but the reported data appear to be internally consistent. On the other hand, McFadden (1975) found magnitude estimation to be unsuitable for obtaining loudness functions for pure tones that varied both in duration and level. His data showed very

large variability and his results were inconsistent with other measures of loudness as a function of duration. It is unclear why the results of these two studies differ so dramatically.

The purpose of the present study is to evaluate the suitability of magnitude estimation as a method for measuring loudness functions for pure tones of different durations. In order to assess this, magnitude estimation data are compared with loudness matches to examine their internal consistency.

### **Procedure**

All stimuli were presented monaurally via headphones. Listeners were seated in a double-wall sound-attenuating booth.

#### **1. Absolute Thresholds**

Absolute thresholds were measured monaurally for 1-kHz tones with 5- and 200-ms durations. Measurements were performed using a three-down, one-up adaptive staircase method and a two-interval, two-alternative, forced-choice paradigm with feedback.

#### **2. Absolute Magnitude Estimation**

Each listener was asked to rate the loudness of individual tones by assigning a number whose magnitude matched the tone's loudness. The number was typed on a computer-microterminal keypad. 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 ranging from 5 dB SL to 100 dB SPL for the 200-ms tones and 110 dB SPL for the 5-ms tones. Ten magnitude estimates were obtained for each level and duration.

The trials were chosen by selecting each new stimulus 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. The final estimates were obtained as the geometric mean of the ten estimates completed for each duration and level.

#### **3. Loudness Matching**

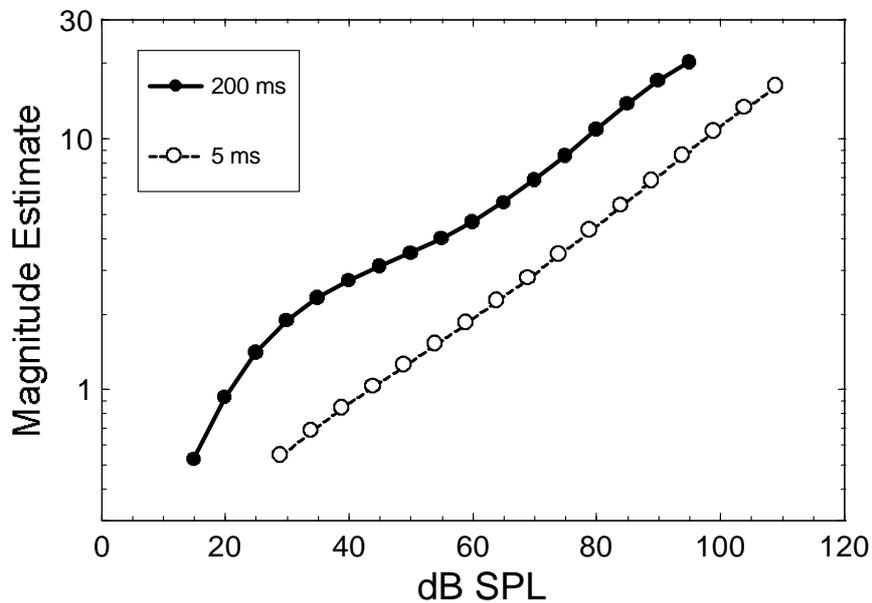
The final part of the experiment consisted of loudness matches between 5- and 200-ms stimuli. This was performed using a roving-level two-alternative forced-choice adaptive procedure similar to that used by Buus *et al.* (1999). This procedure obtains ten concurrent loudness matches by randomly interleaving 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.

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 the response terminal. The level of the variable tone was adjusted according to a simple up-down method. 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. Each track ended after nine reversals and the average level of the last four reversals was used as an estimate of the point of subjective equality. 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).

### Results and Discussion

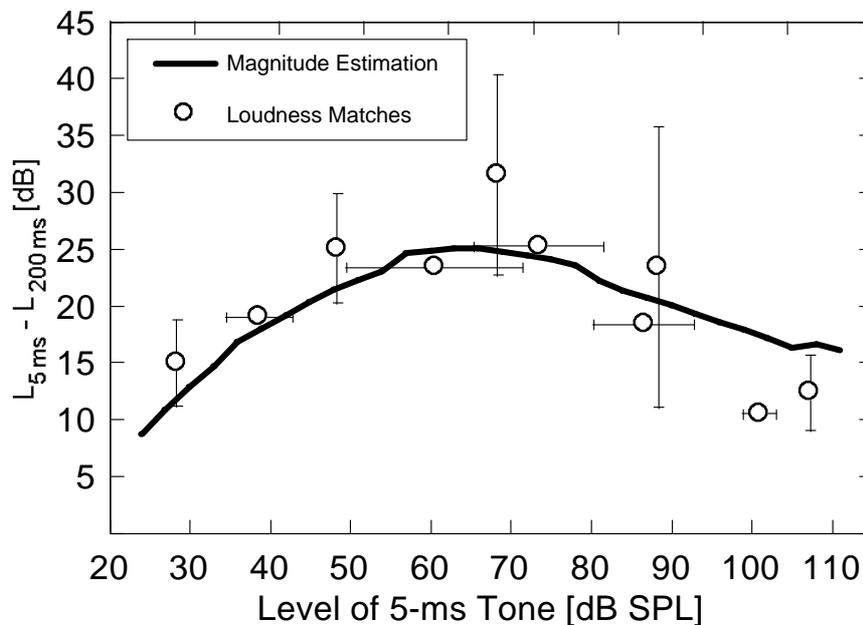
Figure 1 shows the geometric mean of nine listeners' loudness functions obtained using magnitude estimation for long and short tones. The loudness functions have a mid-to-high-level slope of about 0.18. This is lower than the frequently reported slope of 0.3 (Hellman, 1991), but it is within the range of previously reported slopes (cf. Viemeister and Bacon, 1988).



**Figure 1.** Geometric mean of nine listeners' loudness functions for long and short tones obtained with magnitude estimation.

The average amount of temporal integration—defined as the level difference between equally loud short and long tones—is shown in Figure 2. The open circles show the mean loudness-matching results and the line shows the amount of temporal integration derived by averaging level differences between points yielding equal magnitude estimates for each listener. In other words, the line shows the average horizontal distance between the long and short loudness functions.

The temporal-integration functions obtained from loudness matching and magnitude estimation (Fig. 2) agree with one another and are consistent with other similar studies (Florentine *et al.*, 1996; Florentine *et al.*, 1998; Buus *et al.*, 1999) in both magnitude and form. The amount of temporal integration varies non-monotonically with level and is largest at moderate levels.



**Figure 2.** Temporal integration of loudness derived from loudness matches and magnitude estimation; Error bars indicate the standard deviation of the means of 9 subjects.

However, individual data (not shown here due to space constraints) are highly variable. Some listeners exhibit good agreement between loudness matches and magnitude estimation, but others do not. The general form of the loudness functions for all listeners was clear, but detailed information was obscured by variability in most cases. It is noteworthy that some listeners produce results with enough variability to prevent precise details of loudness functions of different durations to be obtained using magnitude estimation, although the average data appear accurate.

Whereas variability of magnitude estimation appears to obscure details of loudness functions in individual listeners, other procedures, such as cross-modality matching using string length (Florentine *et al.*, this volume) seem able to reveal such details. Cross-modality matching appeared to be more successful with the same listeners and comparable instruction and training.

### Conclusion

Magnitude estimation of loudness for tones with different durations produces fairly high variability in individual listeners. Some listeners were able to provide reliable results that were consistent with direct loudness matches, but others were not. Therefore, experimenters should use caution when using magnitude estimation to determine individuals' loudness functions for

pure tones of various durations. Mean data, however, appear to be useful in assessing the shape of the loudness function for tones of different durations.

#### Acknowledgment

This research was supported by NIH/NIDCD grant R01DC02241.

#### References

- 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.
- Florentine, M., Epstein, M., and Buus, S. This volume.
- Hellman, R. P. (1991). Loudness scaling by magnitude scaling: Implications for intensity coding. In: G. A. Gescheider and S. J. Bolanowski (Eds.), *Ratio Scaling of Psychological Magnitude: In Honor of the Memory of S. S. Stevens*, Hillsdale, NJ: Erlbaum.
- Hellman, R. P. and Zwislocki, J. J. (1964). Loudness function of a 1000-cps tone in the presence of a masking noise. *J. Acoust. Soc. Am.*, **36**, 1618-1627.
- Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *J. Acoust. Soc. Am.*, **49**, 467-477.
- McFadden, D. (1975). Duration-intensity reciprocity for equal loudness. *J. Acoust. Soc. Am.*, **57**, 702-704.
- Poulton, E. C. (1989). *Bias in Quantifying Judgments*. Hillsdale, NJ: Erlbaum.
- Stevens, J. C. and Hall, J. W. (1966). Brightness and loudness as a function of stimulus duration. *Perc. Psychophys.*, **1**, 319-327.
- Stevens, S.S. (1975). *Psychophysics*. New York: Wiley.
- Viemeister, N. F. and Bacon, S. P. (1988). Intensity discrimination, increment detection, and magnitude estimation for 1-kHz tones. *J. Acoust. Soc. Am.*, **84**, 172-178.