A test of the Binaural Equal-Loudness-Ratio hypothesis for tones[a]

Jeremy Marozeau[b]
Communication Research Laboratory, Department of Speech-Language Pathology and Audiology (106A FR), Institute for Hearing, Speech and Language, Northeastern University, 360 Huntington Avenue, Boston, Massachusetts 02115

Michael Epstein
Auditory Modeling and Processing Laboratory, Department of Speech-Language Pathology and Audiology (106A FR), Communications and Digital Signal Processing Center, ECE Department (440 DA), Institute for Hearing, Speech, and Language, Northeastern University, 360 Huntington Avenue, Boston, Massachusetts 02115

Mary Florentine and Becky Daley
Communication Research Laboratory, Department of Speech-Language Pathology and Audiology (106A FR), Institute for Hearing, Speech and Language, Northeastern University, 360 Huntington Avenue, Boston, Massachusetts 02115

(Received 4 April 2006; revised 19 September 2006; accepted 21 September 2006)

It is well known that a tone presented binaurally is louder than the same tone presented monaurally. It is less clear how this loudness ratio changes as a function of level. The present experiment was designed to directly test the Binaural Equal-Loudness-Ratio hypothesis (BELRH), which states that the loudness ratio between equal-SPL monaural and binaural tones is independent of SPL. If true, the BELRH implies that monaural and binaural loudness functions are parallel when plotted on a log scale. Cross-modality matches between string length and loudness were used to directly measure binaural and monaural loudness functions for nine normal listeners. Stimuli were 1-kHz 200-ms tones ranging in level from 5 dB SL to 100 dB SPL. A two-way ANOVA showed significant effects of level and mode (binaural or monaural) on loudness, but no interaction between the level and mode. Consequently, no significant variations were found in the binaural-to-monaural loudness ratio across the range of levels tested. This finding supports the BELRH. In addition, the present data were found to closely match loudness functions derived from binaural level differences for equal loudness using the model proposed by Whilby *et al.* [J. Acoust. Soc. Am. **119**, 3931–3939 (2006)]. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2363935]

PACS number(s): 43.66.Cb, 43.66.Pn, 43.66.Ba [AJO] Pages: 3870–3877

I. INTRODUCTION

A tone presented to both ears is perceived as louder than the same tone presented to only one ear (Fletcher and Munson, 1933; Scharf and Fishken, 1970; Marks, 1978). Generally, it is assumed that this ratio is equal to 2 for dichotic tones at the same loudness [for a review, see Marks (1978) and Hellman (1991)], but some studies have suggested a lower ratio (Scharf and Fishken, 1970; Zwicker and Zwicker, 1991; Whilby *et al.*, 2006). The ratio has been assumed to be independent of level, even though it is unclear if this is true (Fletcher and Munson, 1933; see also Fletcher, 1953). This assumption of a constant ratio between the binaural and monaural loudness growth functions is known as the Binaural Equal-Loudness-Ratio hypothesis (BELRH). Whereas Algom *et al.*'s (1989) magnitude estimation of monaural and binaural data lend support to the BELRH for tones, other studies report that the binaural loudness function is steeper than the monaural loudness function for tones (Hellman and Zwislocki, 1963) and for noises (Reynolds and Stevens, 1960), implying a monotonic increase of the ratio as function of level.

Binaural loudness summation can be studied with two different measures: the Binaural Level Difference for Equal Loudness (BLDEL) (Whilby *et al.*, 2006) and the binaural-to-monaural loudness ratio. When the binaural and the monaural loudness functions are presented in the same logarithmic plot, the first measure corresponds to the horizontal distance between the two functions, and the second to the vertical distance. If the BELRH is assumed, then the two functions should be parallel (i.e., the vertical distance is constant). In this case the slope of the loudness function will be proportional to the inverse of the BLDEL. Therefore, it is possible to derive a loudness function directly from the BLDEL data (Fletcher and Munson, 1933; Whilby *et al.*, 2006) using this assumption. On the other hand, without this

---

[a] A portion of this work was presented at the Midwinter Meeting of the Association for Research in Otolaryngology [Marozeau *et al.*, "New method for deriving loudness functions from Binaural Level Differences for Equal Loudness" (2006)] and the American Auditory Society Meeting [Marozeau *et al.*, "Testing the Binaural Equal-Loudness-Ratio hypothesis using Cross-modality Matching" (2006)].

[b] Electronic mail: j.marozeau@neu.edu

assumption, the BLDEL data alone do not provide any information on the shape of the loudness function. Therefore, in order to test the BELRH, binaural and monaural loudness functions must be measured directly over a large range of levels.

Such direct measurements of loudness functions have been previously performed for tones using magnitude estimation (Hellman and Zwislocki, 1963; Scharf and Fishken, 1970; Marks, 1978; Algom *et al.*, 1989). Unfortunately, this procedure may be susceptible to biases, especially at the extremes of the range (Poulton, 1989), and may not be suitable for the measurement of individual loudness functions (Epstein and Florentine, 2006). Therefore, insufficient data are available for accurately assessing the BELRH over a wide range of levels for pure tones in individual listeners. One of the two purposes of the present experiment is to obtain a reliable set of data that is comprehensive enough to permit testing the BELRH for individual listeners.

Obtaining reliable loudness functions for individual listeners requires proper psychophysical procedures. Given the high variability of magnitude-estimation data for loudness in individuals (Epstein and Florentine, 2006), Epstein and Florentine (2005) employed a cross-modality-matching procedure to measure loudness functions of 5- and 200-ms tones. Such cross-modality-matching procedures have been shown to yield reliable data for both groups and individuals (Teghtsoonian and Teghtsoonian, 1983; Hellman and Meiselman, 1988; 1990; 1993; Hellman, 1999). Epstein and Florentine (2005) used cross-modality matching to test the Equal-Loudness-Ratio hypothesis for duration, which states that the loudness ratio between short and long tones is independent of the level. Their data support this hypothesis, except at low levels. The present experiment uses the same protocol to test the BELRH, except for the type of stimuli (i.e., monaural and binaural tones were used instead of short and long tones) and the inclusion of a sequence of mid-to-high-level tones every five trials. This sequence was introduced to better replicate the conditions present in Whilby *et al.* (2006).

Similar conditions between Whilby *et al.* (2006) and the present study are important because the second purpose of the present experiment is to obtain data that can be used to test the loudness model proposed by Whilby *et al.* (2006). The model was derived from loudness matches between monaural and binaural tones using a wide range of stimulus levels that varied randomly within a session. According to Silva and Florentine (2006), such data may be affected by Induced Loudness Reduction (ILR). This effect occurs when a moderate-level tone is preceded by a louder *inducer* tone presented at the same or a nearby frequency [(Marks, 1994; Mapes-Riordan and Yost, 1999; Arieh and Marks, 2003), for a review see Wagner and Scharf (2006)]. The presence of an inducer tone can result in a decrease of the loudness level of the quieter tone by 10 to 15 phons. The amount of ILR can vary quite widely depending on the level, presentation mode, sequence, and duration of the preceding sounds (Mapes-Riordan and Yost, 1999; Nieder *et al.*, 2003). Because ILR most profoundly affects the perception of sounds at moderate levels, it can change the shape of the loudness function by making the slope shallower in the mid-range. The effects of loud inducer tones are time dependent and can last at least 10 min (Arieh and Marks, 2003; Epstein and Gifford, 2006). In order to reduce potential variability due to the context dependence of ILR, the present experiment was designed to maintain similar ILR-producing conditions for each trial throughout the entire experiment by exposing the listener to mid-to-high level tones at regular intervals. Although this design should reduce variability caused by different sequential contexts, it may also result in shallower slopes in the mid range of the measured loudness functions and smaller binaural-to-monaural loudness ratios than seen by other experimenters who utilize procedures that do not consider the sequentially evoked variability of the effects of ILR.

II. METHOD

A. Stimuli

The stimuli were 1-kHz tones with equivalent rectangular durations of 200 ms. The tones had 6.67-ms raised-cosine rises and falls and 195-ms steady-state portions to ensure that almost all the energy of the tone bursts was contained within the 160-Hz-wide critical band centered at 1 kHz (Scharf and Fishken, 1970; Zwicker and Fastl, 1990). Tones were presented both monaurally and binaurally. In the monaural condition, tones were presented to the ear with the higher threshold. In the binaural condition, tones were presented at approximately the same loudness to each ear. Because the threshold microstructure of the two ears differs, sounds presented at the same SL to each ear are closer in loudness than those presented at the same SPL, at least at low-to-moderate levels (Mauermann *et al.*, 2004). Therefore, an attenuation corresponding to the threshold difference between the two ears was applied to tones sent to the ear with the lower threshold to achieve binaural presentations at the same SL in each ear. The tone levels in the ear with the higher threshold ranged from the first multiple of 5 dB above threshold to 100 dB SPL in 5-dB steps.

Because the level presented to the ear with the lower threshold was always less than or equal to the level presented to the opposite ear, the total energy presented in the binaural condition was always less than or equal to two times the energy presented in the monaural condition. If, instead, tones presented to the higher threshold ear were amplified to match the opposite ear in SL, the total energy in the binaural condition could have exceeded two times the monaural energy. In that case, the loudness summation observed could have been attributed to a simple increase in 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), low-pass filtered (TDT FT5, $fc=20$ kHz, 135 dB/oct), attenuated again (TDT PA4), and led to a headphone amplifier (TDT HB6), which fed the earphone of the Sony MDR-V6 headset over the ear with the higher threshold. A third attenuator (TDT PA4) was used before the other headphone amplifier and earphone. This setup ensured that the attenuators could control the stimulus level linearly over at least a 120-dB range and the signal was set to the same SL at both ears. 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 was divided into two parts: measurements of absolute threshold and cross-modality matches.

1. Absolute thresholds

Absolute thresholds were measured separately for each ear at 1 kHz using a two-interval, two-alternative forced-choice paradigm with feedback. Each trial contained two observation intervals, which were marked by lights and separated by 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 indicate which interval contained the signal by pressing a key on a small computer terminal. 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.

Each 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, a track was selected at random from 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 last two reversals and the average of the three tracks was considered the absolute threshold.

2. Cross-modality matching

Listeners were given a virtually unbounded ball of very thin but strong string (embroidery floss), scissors, adhesive tape, and a notebook. Each trial was presented in the middle of a 250-ms visually marked interval. They were asked to cut a piece of string that was as long as the sound was loud. After each presentation, the listeners taped the string segment into a notebook, 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. In order to reduce any task-related biases, embroidery floss was selected because it is very thin and inelastic and can be cut very short. Listeners were encouraged to cut the string as short or as long as they wanted.

All levels and modes (monaural and binaural) were presented in random order. One block of trials contained three trials in each mode at each level. Two blocks of trials were completed such that six cross-modality matches were made for each level and mode. Each block lasted approximately 40 min and blocks were separated by a 15-min break. On each trial, a new tone level and mode were randomly selected from all other stimuli that had not yet been presented three times and had a level within 30 dB of the level of the previous trial. If no stimuli fulfilled these criteria, but some other stimuli still had been presented fewer than three times, a dummy trial was inserted. The dummy trial had the same mode and a level 30 dB above or below the preceding level, depending on the levels of the stimuli that remained to be presented. Large level differences between trials were avoided in order to prevent startling the listener from a sudden level increase or a missed stimulus from a sudden level decrease. The dummy trials were included in the final analysis.

As described earlier, this experiment was also designed to normalize the influence of ILR on listener responses. In order to achieve this, a random sequence of stimuli without repetition was presented to the listeners every five trials. The sequence was composed of all of the stimuli ranging from 50 to 100 dB SPL in 5-dB steps in both modes. Before each session, six such sequences were presented in a row to ensure that the same amount of ILR would affect all trials. The final cross-modality matches were the geometric means of the string lengths that were cut to match a given level.

D. Listeners

As shown in Table I, nine listeners (six females and three males) participated in this experiment. They ranged in age from 22 to 74 years old with a mean of 32 years. All listeners had audiometric thresholds less than 20 dB HL (ANSI, 2004) at octave frequencies from 250 to 8000 Hz in both ears when measured clinically and medical histories were consistent with normal hearing. The mean absolute threshold difference between the two ears was 3.7 dB at 1 kHz; no listener had a difference greater than 8 dB. Only L3 (the first author) and L5 had previous experience making loudness judgments.

E. Data analysis

Each data point shows the geometric mean of at least six string lengths. The standard deviation was determined from the logarithms of the string lengths. The group mean and standard error were calculated across the individual listeners' geometric means for each mode and level. The resulting data were transformed back into the string-length domain to show the probable range of each individual listener's responses. To compute average matches across listeners, string lengths for each individual were normalized by dividing by the geometric mean of that individual's string lengths for a binaural tone set to 40 dB SPL.

To examine the effects of stimulus variables, a two-way analysis of variance (ANOVA) for repeated measures was performed on the logarithms of the string lengths (level in dB SPL × presentation mode). The outcome was considered significant when $p \leq 0.05$.

TABLE I. Summary of the individual data (gender, age, thresholds at 1 kHz, and results). ΔL represents the absolute value of the threshold difference between the two ears of each listener. The Bi/Mono ratio is the averaged difference between the binaural and monaural polynomial fits to the data; Monaural exponent and Binaural exponent show the values of the exponents of the fitted power functions for levels above 40 dB SPL.

		Threshold measures in dB HL			Results			
Listener	Gender	Age	Left ear	Right ear	ΔL	Bi/Mono ratio	Monaural exponent	Binaural exponent
L1	F	44	13	7	6	1.24	0.17	0.18
L2	F	24	−4	−1	3	1.3	0.07	0.1
L3	M	30	−4	−4	0	1.4	0.14	0.16
L4	M	25	6	8	2	1.15	0.12	0.14
L5	F	22	3	−2	5	1.38	0.21	0.21
L6	F	74	−4	−9	5	-	-	-
L7	F	28	0	−8	8	1.33	0.15	0.14
L8	M	23	10	8	2	1.28	0.13	0.15
L9	F	22	4	6	2	1.25	0.14	0.15
Average		32	2.6	0.5	3.7	1.29	0.14	0.15

III. RESULTS

The geometric mean of string length is plotted on a log scale as a function of level in Fig. 1 for each individual listener. The full range of string lengths cut was from 0.1 to 36.6 cm with an average of 3.05 cm for the monaural and 4.24 cm for the binaural. The data for individual listeners were generally consistent, as indicated by the small standard deviations and the general monotonicity.

Third-order polynomials were used to estimate the monaural and binaural loudness functions for each listener using a least-squares fit. As is typical of loudness growth functions, there were clear differences among listeners. However, except for two listeners (L6 and L9 at low levels), the binaural fit was always higher than the monaural fit for any given level. This result implies that at an equal level, binaural tones were perceived as louder than monaural tones. Data from L6 in one session were corrupted because she taped the pieces of string on the wrong pages and skipped a page. Therefore, L6's data were not included in the average plot or statistical analysis. The binaural fit for L9 was lower than the monaural fit at low levels. This discrepancy can be attributed to the poor reliability of this listener's responses (i.e., large standard deviations) at those levels. However, except for low levels, the overall shapes of L9's loudness functions seem in agreement with those of the rest of the listeners, so L9's data were included in all subsequent analysis.

For most of the listeners, the cross-modality matching functions for tones were shallower at moderate levels than at low and high levels. They were 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 between the string lengths matched to equal-SPL monaural and binaural tones. Visually, it is approximately independent of SPL for most of the listeners, though some exceptions are apparent.

The mean ratios between the binaural and monaural loudness functions are summarized in Table I. These values ranged from 1.15 to 1.38. The range of ratios measured under very different conditions and found in other studies is about 1.5 to 2.0 (Zwicker and Zwicker, 1991; Marks, 1978).

The relationship between loudness and intensity is classically modeled as a power function of the intensity with an exponent of 0.3 (Stevens, 1955). However, this relationship is more complex than just a straight line on a log scale (Florentine and Epstein, 2006). Near threshold, the relationship between loudness and intensity can be approximated by a power function with an exponent of unity (Zwislocki, 1965; Buus *et al.*, 1998). At moderate levels, the slope of the loudness function is shallower than usually assumed (Florentine *et al.*, 1996), and the function can be locally approximated as a power function with an exponent of 0.2 (Buus and Florentine, 2001). However, in order to compare the present data with classical data in the literature, data from this experiment were fitted with a simple power function for all levels above 40 dB SPL. Table I shows the values of the exponents for individual listeners. These values are lower than the classical power function exponent of 0.3, but consistent with some other studies, as discussed later in Sec. IV A.

The group mean data and standard errors are plotted in Fig. 2. They are plotted in the same manner as Fig. 1, but the *Y* axis is the normalized string length. The average data show the same general trends as the majority of the individual data. Both loudness functions were shallower at moderate levels than at low and high levels. The thick line showing the binaural-to-monaural loudness ratio is nearly horizontal, indicating that it is approximately independent of SPL. This observation is supported by an ANOVA. The effects of the level and mode are both highly significant ($p < 0.0001$), but the interaction between them is not ($p = 0.12$), as is expected if the effect of mode is independent of the level. Accordingly, the results of the ANOVA support the BELRH.

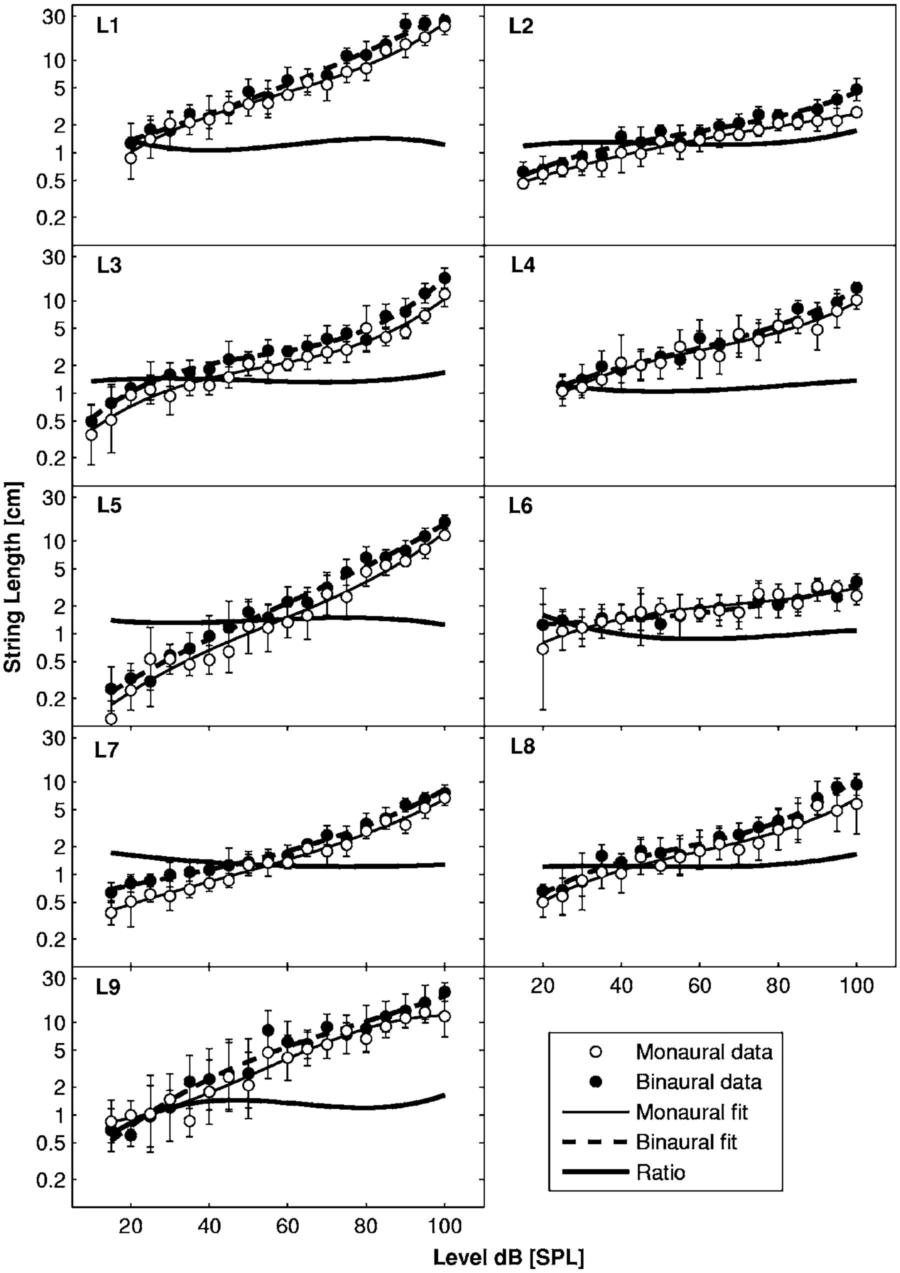

FIG. 1. Individual loudness functions obtained from all nine listeners. The geometric mean of the string length is plotted on a log scale as a function of level. The filled circles show data for the binaural tones and the open circles show the data for the monaural tones. The vertical bars show ± one standard deviation of the log of the string lengths. The solid thin lines show third-order polynomials fitted to the monaural average data and the dashed lines show third-order polynomials fitted to the binaural average data. The thick lines show the ratio of string lengths obtained for equal-SPL monaural and binaural tones as estimated from the polynomials.

IV. DISCUSSION

A. Comparison with data in the literature

1. Comparison with Epstein and Florentine (2005)

The present experimental protocol differed from that of Epstein and Florentine (2005) by only the type of stimuli used (i.e., monaural and binaural tones rather than 5- and 200-ms tones). Both experiments used monaural 200-ms tones at 1 kHz over a wide range of levels. Therefore, it is possible to directly compare the average loudness functions extracted from the two experiments. In order to be consistent with Epstein and Florentine (2005), the group means of the string lengths for monaural tones above 40 dB SPL were approximated with a power function. The value of the exponent fit to the data was 0.14 for both studies, indicating that the average loudness functions in the present experiment grew at nearly the same rate as the loudness functions measured by Epstein and Florentine (2005). Figure 3 shows the average data from the two experiments on the same graph. To facilitate comparison, the two loudness functions were normalized so that the match to the 40 dB SPL tone had a value of one. The two functions are quite similar, showing that the difference in methodology between the two experiments (i.e., the inclusion of a sequence of mid-to-high level tones) did not affect the slope of the loudness function for monaural tones.

2. Comparison with classical loudness experiments

Although the exponent of 0.14 found in the present study is quite similar to that found by Epstein and Florentine (2005), this value and the binaural-to-monaural loudness ratio of 1.29 are smaller than the values found in most prior studies using different experimental paradigms. It may be possible that judgments were influenced by the reluctance of listeners to cut long pieces of string, and the difficulty of

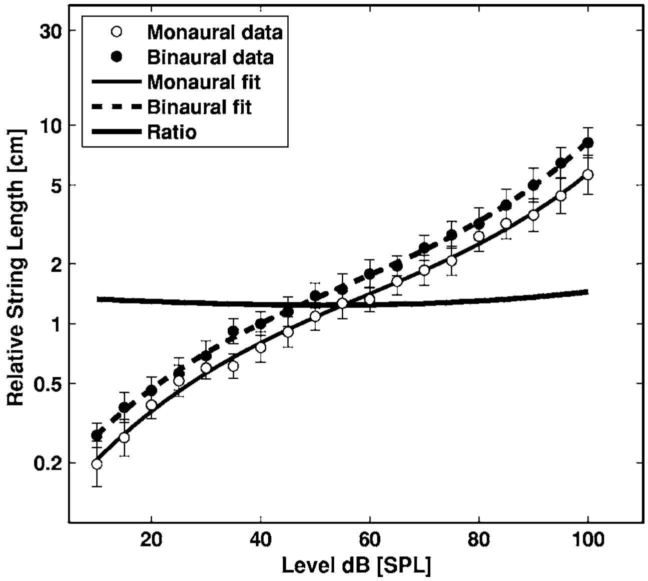

FIG. 2. Average data from the individual listeners (L6 omitted). The third-order polynomials fitted to the data were normalized by dividing each length by the average string length for the 40 dB SPL tone in the binaural condition. The vertical bars show ± one standard error of the log of the string length.

cutting very short pieces of string. Although this possibility cannot be completely ruled out, the data do not seem to show these effects (i.e., no unexpected compression appears at the very low and high levels). Most of these studies used magnitude estimation tasks to obtain loudness slopes and binaural-to-monaural ratios. Work by Treisman and Irwin (1967) indicates that different tasks may involve different perceptual processes and yield different results (i.e., the transformation of the psychological loudness scale to numbers can differ from the transformation to string lengths). Prior experimenters have made corrections to compensate for differences in methodology. For example, Marks (1987) raised each numerical estimate to the power of 0.3/0.2 to adjust the slope of his loudness function to the expected 0.3. If the current data were adjusted to match the slope of a fitted power function to 0.3, then the binaural-to-monaural loudness ratio would become 1.7. This ratio is well within the range of ratios found by other experimenters (Scharf and Fishken, 1970; Zwicker and Zwicker, 1991). This transformation suggests that the slope of the loudness function and the binaural-to-monaural loudness ratio are dependent on each other, and that both are dependent on the methodology.

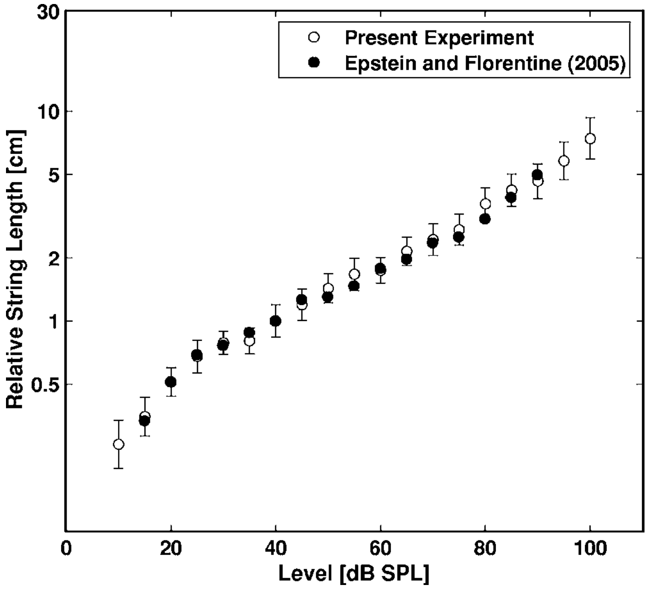

FIG. 3. Comparison between the average loudnesses of the monaural 200-ms tones obtained in the present paper (open circles) and the average loudnesses of the 200-ms tones (filled circles) obtained in Epstein and Florentine (2005).

B. Testing the Binaural Equal-Loudness-Ratio hypothesis

The present data clearly support the BELRH and are consistent with other studies that used magnitude estimation to measure the ratio (Scharf and Fishken, 1970; Marks, 1987; Algom *et al.*, 1989; Zwicker and Zwicker, 1991). Because the binaural and monaural loudness functions maintain a constant vertical distance on a log scale (as shown in Figs. 1 and 2) and the slopes of both functions change at moderate levels, the level difference between binaural and monaural tones required for equal loudness changes as a function of level. According to the present data, the Binaural Level Difference for Equal Loudness (BLDEL) is 5 dB at 20 dB SPL, 8 dB at 66 dB SPL, and 7 dB at 100 dB SPL.

C. The effect of induced loudness reduction in loudness judgments

The sequences of mid-to-high-level tones were introduced to better match the experimental conditions of Whilby *et al.* (2006) because it was hypothesized that Induced Loudness Reduction (ILR) affected those data. ILR is known to affect sounds differently as a function of level (Mapes-Riordan and Yost, 1999; Nieder *et al.*, 2003). Sounds below 40 dB SPL are not susceptible to ILR. Above 40 dB SPL the amount of ILR increases with level, at least up to 70 dB SPL (Scharf *et al.*, 2005). Consequently, loudness functions derived from data that are influenced by ILR are likely to have a shallower slope at moderate levels. The results, shown in Fig. 3, do not indicate that the amount of ILR produced by the sequence of tones used in the present experiment is greater than the amount of ILR seen due to the simple randomization of stimulus levels in Epstein and Florentine (2005). This is consistent with measures showing the rapidity of onset and the duration of the effects of ILR (Epstein and Gifford, 2006).

D. Modeling growth of loudness for individual listeners

Whilby *et al.* (2006) proposed a method for deriving loudness functions from BLDEL data. Briefly, the monaural loudness function, F_m, was defined as

$$F_m(L) = 10^{aL^3+bL^2+cL+d_m}, \tag{1}$$

where L was the level in dB SPL of the monaural stimulus and a, b, c, and d_m were the free parameters of the model. Because of the assumption of BELRH, the binaural loudness

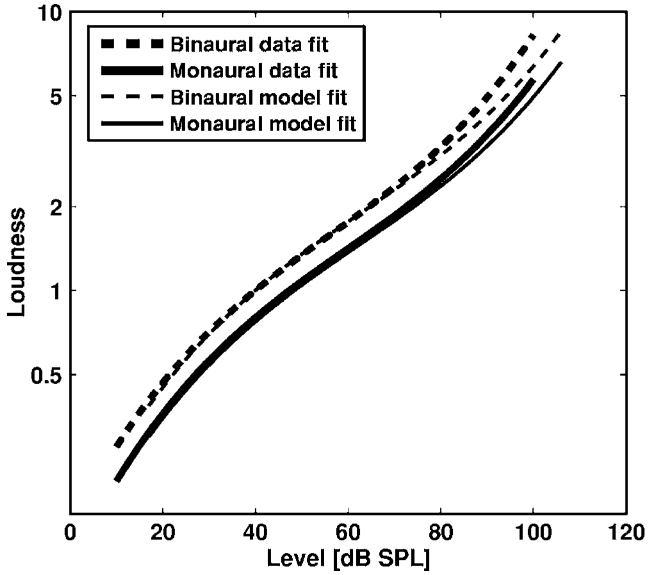

FIG. 4. Comparison between binaural (dashed lines) and monaural (solid lines) loudness functions extracted in the present paper (thick lines) and the loudness functions derived from the model (thin lines) Whilby *et al.* (2006).

function, F_b, *wa*s assumed to be proportional to F_m:

$$F_b(L) = \gamma F_m(L) = 10^{aL^3+bL^2+cL+d_b}, \tag{2}$$

where γ was the binaural-to-monaural loudness ratio and was equivalent to $10^{d_b-d_m}$. Because d_b can be set to normalize the scale and γ can be set at the value found in the present experiment, only three free parameters need to be determined to derive the two loudness functions. For a fixed selected level of the monaural stimulus, L_m, the BLDEL data from Whilby *et al.* (2006) were used to estimate the level of the binaural stimulus, L_b, at which the loudnesses of the monaural and binaural stimuli were equal. This can be expressed by the following equations:

$$F_b(L_b) = F_m(L_m),$$

$$\log\,(F_b(L_b)) = \log\,(F_m(L_m)), \tag{3}$$

$$a(L_m^3 - L_b^3) + b(L_m^2 - L_b^2) + c(L_m - L_b) = d_b - d_m = \log\,(\gamma).$$

First, the BLDEL data for each listener were fitted with a third-order polynomial. Then, by using a least-squares fit, the three free parameters (a, b, and c) were selected to minimize the error of the fit between the model and the BLDEL data.

Figure 4 shows the monaural and binaural loudness functions derived using the model compared with the functions extracted from the present experiment. The model-derived functions and the measured functions are quite similar up to 80 dB SPL. Above 80 dB SPL, the functions derived using the model are shallower than the experimentally obtained functions. These results show support for the model, at least up to 80 dB SPL.

This model may be valuable for predicting the shape of a loudness function for a given binaural-to-monaural loudness ratio. Although it does require experimental data, it has some advantages over the direct measurement of a loudness function using a task like magnitude estimation or cross-modality matching. First, the experimental data needed to derive loudness functions using the model are collected using a simple, reliable task; listeners adjust the level of the monaural tone to match the loudness of the binaural tone, and *vice versa*. Unfortunately, the variability of the two types of tasks cannot be easily compared because they involve different scales (i.e., string lengths versus dB SPL). However, it is worth noting that there is less difference between each individual loudness function when derived using the model than when extracted directly in the present experiment. Second, the present cross-modality matching results show great similarity to the data of Epstein and Florentine (2005). This suggests that a task like magnitude estimation or cross-modality matching, which involves the presentation of a wide range of levels in random order, causes the same amount of ILR as a sequence of tones chosen specifically to produce ILR. It is possible to present levels in ascending order to ensure that a listener never hears a high-level tone that precedes a moderate-level tone. Unfortunately, this may introduce another type of bias in which the listener will be able to anticipate the loudness of the upcoming tone presentation. It then seems difficult to design an experiment that can directly extract a loudness function that is unaffected by ILR. However, Silva and Florentine's (2006) data suggest that using an adaptive psychophysical loudness-matching procedure that presents levels in an ascending order can reduce the effects of ILR. Therefore, it may be possible to derive loudness functions that may not be affected by ILR from BLDEL data collected using this type of experimental design.

V. CONCLUSIONS

For most of the individual listeners, binaural and monaural loudness functions are parallel when plotted on a log scale. This result is even more clear in the average data. An ANOVA did not find a significant effect of level on the binaural-to-monaural loudness ratio and, thus, the ratio can be considered to be independent of level. This finding supports the Binaural Equal-Loudness-Ratio hypothesis (BELRH). A comparison of the present experiment with that of Epstein and Florentine (2005) suggests that a simple, random presentation of tones over a wide range of levels, as used in a typical magnitude estimation or cross-modality matching experiment, is sufficient to cause as much Induced Loudness Reduction (ILR) as a sequence of frequently presented mid-to-high level tones. The ratio between the average binaural and monaural loudness functions was found to be lower than the value often reported. This suggests that the binaural loudness summation for tones is less than perfect (i.e., the ratio is less than 2). Finally, the finding of no significant effect of level on the binaural-to-monaural loudness ratio lends support to the model proposed by Whilby *et al.* (2006) for deriving loudness functions from Binaural Level Difference for Equal Loudness (BLDEL) data.

ACKNOWLEDGMENTS

The authors wish to thank Andrew Oxenham, Brian Moore, and an anonymous reviewer for their helpful comments. This research was supported by NIH/NIDCD Grant No. R01DC02241.

Algom, D., Rubin, A., and Cohen-Raz, L. (**1989**). "Binaural and temporal integration of the loudness of tones and noises," Percept. Psychophys. **46**, 155–166.

American National Standards Institute (**2004**). "Specification for Audiometers (ANSI S3.6-2004)," (ANSI, New York).

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

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), 236–241.

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

Epstein, M., and Florentine, M. (**2005**). "A test of the Equal-Loudness-Ratio hypothesis using cross-modality matching functions," J. Acoust. Soc. Am. **118**, 907–913.

Epstein, M., and Florentine, M. (**2006**). "Loudness of brief tones measured by magnitude estimation and loudness matching," J. Acoust. Soc. Am. **119**, 1943–1945.

Epstein, M., and Gifford, E. (**2006**). "A potential carry-over effect in the measurement of induced loudness reduction," J. Acoust. Soc. Am. **120**, 305–309.

Fletcher, H. (**1953**). *Speech and Hearing in Communication* (Van Nostrand, New York).

Fletcher, H., and Munson, W. A. (**1933**). "Loudness, its definition, measurement and calculation," J. Acoust. Soc. Am. **5**, 82–108.

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., and Epstein, M. (**2006**). "To honor Sevens and repeal his law (for the auditory system)," in *Fechner Day 2006*, edited by D. E. Kornbrot, R. M. Msetfi, and A. W. MacRae (St. Albans, England), 37–42.

Hellman, R. (**1991**). "Loudness measurements by magnitude scaling: Implications for intensity coding," in *Ratio Scaling of Psychological Magnitude* edited by S. J. Bolanowski and G. A. Gescheider (Erlbaum, Hillsdale, NJ), pp. 215–227.

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. (**1963**). "Monaural loudness function at 1000 cps and interaural summation," J. Acoust. Soc. Am. **35**, 856–865.

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

Mapes-Riordan, D., and Yost, W. A. (**1999**). "Loudness recalibration as a function of level," J. Acoust. Soc. Am. **106**, 3506–3511.

Marks, L. E. (**1978**). "Binaural summation of the loudness of pure tones," J. Acoust. Soc. Am. **64**, 107–113.

Marks, L. E. (**1987**). "Binaural versus monaural loudness: Supersummation of tone partially masked by noise," J. Acoust. Soc. Am. **81**, 122–128.

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

Mauermann, M., Long, G. R., and Kollmeier, B., (**2004**). "Fine structure of hearing threshold and loudness perception," **116**, 1066–1080.

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.

Poulton, E. C. (**1989**). *Bias in Quantifying Judgments* (Erlbaum, Hillsdale, NJ).

Reynolds, G. S., and Stevens, S. S. (**1960**). "Binaural summation of loudness," J. Acoust. Soc. Am. **32**, 1337–1344.

Scharf, B., and Fishken, D. (**1970**). "Binaural summation of loudness: reconsidered," J. Exp. Psychol. **86**, 374–379.

Scharf, B., Wagner, E., and Nieder, B. (**2005**). "Induced loudness reduction: A review," J. Acoust. Soc. Am. **117**, 2454.

Silva, I., and Florentine, M. (**2006**). "Effect of adaptive psychophysical procedure on loudness matches," J. Acoust. Soc. Am. **120**, 2124–2131.

Stevens, S. S. (**1955**). "The measurement of loudness," J. Acoust. Soc. Am. **27**, 815–829.

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

Treisman, M., and Irwin, R. J. (**1967**). "Auditory intensity discriminal scale I. Evidence derived from binaural intensity summation," J. Acoust. Soc. Am. **42**, 586–592.

Wagner, E., and Scharf, B. (**2006**). "Induced loudness reduction as a function of exposure time and signal frequency," J. Acoust. Soc. Am. **119**, 1012–1020.

Whilby, S., Florentine, M., Wagner, E., and Marozeau, J. (**2006**). "Monaural and binaural loudness of 5- and 200-ms tones in normal and impaired hearing," J. Acoust. Soc. Am. **119**, 3931–3939.

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

Zwicker, E., and Zwicker, U. T. (**1991**). "Dependence of binaural loudness summation on interaural level differences, spectral distribution, and temporal distribution," J. Acoust. Soc. Am. **89**, 756–764.

Zwislocki, J. J. (**1965**). "Analysis of some auditory characteristics," in *Handbook of Mathematical Psychology*, edited by R. D. Luce, R. R. Bush, and E. Galanter (Wiley, New York).