

Binaural loudness summation for speech presented via earphones and loudspeaker with and without visual cues^{a)}

Michael Epstein^{b)}

Auditory Modeling and Processing Laboratory, Department of Speech-Language Pathology and Audiology (106A FR), Communications and Digital Signal Processing Center, Department of Electrical and Computer Engineering, Northeastern University, 360 Huntington Avenue, Boston, Massachusetts 02115

Mary Florentine

Communication Research Laboratory, Department of Speech-Language Pathology and Audiology (106A FR), and Department of Electrical and Computing Engineering, Northeastern University, 360 Huntington Avenue, Boston, Massachusetts 02115

(Received 19 May 2011; revised 21 March 2012; accepted 21 March 2012)

Preliminary data [M. Epstein and M. Florentine, *Ear. Hear.* **30**, 234–237 (2009)] obtained using speech stimuli from a visually present talker heard via loudspeakers in a sound-attenuating chamber indicate little difference in loudness when listening with one or two ears (i.e., significantly reduced binaural loudness summation, BLS), which is known as “binaural loudness constancy.” These data challenge current understanding drawn from laboratory measurements that indicate a tone presented binaurally is louder than the same tone presented monaurally. Twelve normal listeners were presented recorded spondees, monaurally and binaurally across a wide range of levels via earphones and a loudspeaker with and without visual cues. Statistical analyses of binaural-to-monaural ratios of magnitude estimates indicate that the amount of BLS is significantly less for speech presented via a loudspeaker with visual cues than for stimuli with any other combination of test parameters (i.e., speech presented via earphones or a loudspeaker without visual cues, and speech presented via earphones with visual cues). These results indicate that the loudness of a visually present talker in daily environments is little affected by switching between binaural and monaural listening. This supports the phenomenon of binaural loudness constancy and underscores the importance of ecological validity in loudness research. © 2012 Acoustical Society of America. [http://dx.doi.org/10.1121/1.3701984]

PACS number(s): 43.66.Cb, 43.66.Pn, 43.66.Ba, 43.66.Lj [CJP]

Pages: 3981–3988

I. INTRODUCTION

Are conclusions about loudness drawn from tones presented via earphones in laboratories applicable to listening to a talker in a room? Laboratory experiments using earphones indicate that a tone presented binaurally is louder than the same tone presented monaurally (Fletcher and Munson, 1933). This phenomenon is known as binaural loudness summation, BLS. Based on earlier work, it had been generally assumed that the binaural-to-monaural loudness ratio is equal to two for diotic or dichotic tones at the same loudness (for a review, see Hellman, 1991; Marks, 1978; Sivonen and Ellermeier, 2011). In other words, a tone presented to two ears is twice as loud as a tone presented to only one ear. More recent studies suggest a lower ratio ranging from 1.29–1.7 (Algom *et al.*, 1989; Epstein and Florentine, 2009; Marozeau *et al.*, 2006; Scharf and Fishken, 1970; Zwicker and Zwicker, 1991). Data from alternate binaural equal loudness matches of tones or noisebands presented via earphones provide additional support for a lower

binaural-to-monaural ratio (Edmonds and Culling, 2009; Scharf, 1969; Whilby *et al.*, 2006). The binaural gain for directional narrow-band noises heard in an anechoic chamber is about 3 dB, which corresponds to an average binaural-to-monaural loudness ratio of 1.2 according to the data and model of Sivonen and Ellermeier (2008) when it is assumed that loudness grows as a power function of intensity with an exponent of 0.3.

When listening in an anechoic sound field, the level and spectrum of sound reaching the tympanic membrane changes as a function of the sound’s location in relation to the person’s orientation. This acoustic pattern is known as the head-related transfer function (HRTF), which depends primarily on the physical properties of the listener’s head and torso. Many experiments have been performed in anechoic sound fields because it is possible to have more control over stimulus presentation. When stimulus presentation in experimental conditions approximate daily environments, such as in reverberant rooms with sighted listeners, the stimulus cues become more complex and interesting phenomena emerge. For example, a type of perceptual constancy occurs in which loudness remains relatively constant while sound source distance is varied (i.e., loudness constancy, see Zahorik and Wightman, 2001; Mohrmann, 1939). In fact, many factors influence the perception of loudness in daily environments. Auditory cues can differ widely depending on the stimulus,

^{a)}A portion of this work was presented as an invited talk to the International Congress of Acoustics [Florentine, M. and Epstein, M. (2010): “Ecological loudness: Binaural loudness constancy”].

^{b)}Author to whom correspondence should be addressed. Electronic mail: m.epstein@neu.edu

context, and mode of presentation (i.e., earphones and free, diffuse, and directional sound fields). For example, sound level and reverberation have different effects on the perceptions of loudness and source distance for speech and non-speech sounds (Warren, 1973). For speech sounds, vocal effort may be weighted more heavily in judgments of loudness than sound level (Brungart and Scott, 2001).

Recent data obtained using speech stimuli heard via loudspeakers from a visually present talker suggest another type of perceptual constancy that involves loudness. These data challenge conclusions drawn from classical measurements obtained in laboratories using earphones. Epstein and Florentine (2009) measured BLS for speech and tones presented via earphones and loudspeakers with a fixed distance from the listener. Their data show (1) the amount of BLS is significantly smaller for speech from a visually present talker than for recorded speech and tones, (2) the amount of BLS is significantly smaller for loudspeaker presentation than for earphone presentation, and (3) the amount of BLS is smallest for speech from a visually present talker presented via loudspeakers than any of their other test conditions. Their data suggest the presence of binaural loudness constancy (i.e., speech from a visually present talker heard under ecologically valid conditions is not much louder when listening with two ears than listening with only one ear).

Because Epstein and Florentine's (2009) experimental design was an initial approach to studying this complex problem, two possible confounding variables were not taken into account. First, they compared conditions in which a visually present talker spoke monitored live voice (MLV) spondees with conditions in which standardized recorded spondees were presented via earphones. There was no assurance that there were no differences between the female MLV talker and the male talker on the recordings. Second, they used two loudspeakers to simultaneously present their stimuli with each at a 45° angle from the listener. It is possible that the use of two speakers could have caused frequency-specific interference that contributed to the less-than-expected BLS. These confounding variables could have been responsible for a possible discrepancy with Cox and Gray's (2001) data. Their data suggest that, even without a visually present talker, speech presented via a single loudspeaker results in less BLS than speech presented via earphones. This prompts the question whether a visually present talker is necessary for binaural loudness constancy.

The present experiment provides a controlled test of the following hypothesis: Speech combined with visual cues from the *same* talkers presented under more ecologically valid conditions with a *single* loudspeaker results in less BLS than speech presented via earphones and/or without visual cues. In addition, it provides a much more comprehensive set of data from 12 listeners, who performed a total of 23 040 loudness judgments.

II. METHOD

A. Stimuli

A set of 16 speech stimuli was recorded with video combined with audio consisting of 4 native talkers (2 males, 2 females) speaking 4 two-syllable words that had equal stress on both syllables (i.e., spondees). Recordings of the stimuli

were made in a quiet room within a single-walled, sound-attenuating booth (Acoustic Systems, Model RE-147S). A head-worn cardioid microphone (AKG model C520) was placed approximately 20 cm from the talker's lips. Microphone signals were amplified using a Mackie VLZ3 mixer/preamp and recorded using Audacity software. The four words (i.e., hothouse, northwest, playground, and woodwork) were repeated several times with moderate vocal effort; the subjectively clearest and most evenly stressed presentation of each word for each talker was selected. The visual stimuli for the video conditions showed a direct, frontal view of the talker's head. The goal of the video display was to create a natural visual distance.

The level of each word was adjusted to match the desired stimulus level by amplifying or attenuating the signal until the root-mean-square, rms, level matched the desired level. Note that this procedure does not necessarily result in equal loudness. Each of the speech stimuli was presented at five rms levels: 25, 40, 55, 70, and 80 dB sound pressure level (SPL).

B. Procedure and apparatus

Each listener estimated loudness using a magnitude estimation paradigm by assigning a number whose magnitude matched the loudness of each stimulus. No reference or range was given as a basis for this judgment. (For procedure and instructional details for magnitude estimation, see Marks and Florentine, 2011.) Stimuli were presented under four conditions: Video combined with audio presented via earphones; only audio presented via earphones; only audio presented via a loudspeaker; and video combined with audio presented via a loudspeaker. Trials were split into eight blocks, two blocks for each condition, one of which was monaurally presented, the other one binaurally presented. Each of these eight blocks was presented three times resulting in 24 total blocks. These 24 blocks were presented to each listener in a separate random order. Each presentation block contained 80 trials: Four words spoken by four talkers presented at five levels. The stimuli in each block were presented in random order. This resulted in 1920 total speech-stimuli trials per listener. The entire experiment took approximately 12 h per listener, performed over multiple days. Listeners were given frequent breaks to prevent fatigue.

For the "monaural" loudspeaker condition, a foam earplug (E.A.R. classic-29 NRR) was inserted and checked for proper fit by the experimenters throughout the experiment. Half of the listeners plugged their left ear, half their right ear for the monaural condition. As no listeners had an ear asymmetry greater than 5 dB at threshold and most of the presentation levels were well above threshold, it is not expected that the choice of plugged ear should have a significant impact on loudness. Earplugs increased threshold by at least 20–24 dB at 1 kHz compared to unoccluded ears, as determined by previous measurements (see Epstein and Florentine, 2009). Although it is impossible to obtain a completely monaural condition in listeners with normal hearing in both ears, the attenuation provided by a monaural earplug was deemed to be sufficient for the present experiment based on data obtained by Zwicker and Zwicker (1991). They made

measurements of BLS with a 20-dB difference between the ears of their listeners; this resulted in a 20% or smaller reduction in the ratio from the true monaural condition in the range presently used. In any case, exact details of the amount of attenuation of the earplug are not critical to this experiment because the attenuation can be assumed to be the same across all loudspeaker conditions; the measurements of primary interest are the relative differences in BLS across these conditions.

The stimuli were presented in MATLAB (2007b running on Windows XP) and converted from digital (48-kHz sampling frequency) to analog using a 24-bit Lynx Two Soundcard. The analog signal was then presented either via Sony MDR-V6 earphones or a single M-Audio BX8a 8-inch studio monitor speaker, 1 m directly in front of the listener in a double-walled test booth (Industrial Acoustics Company). The inner room was approximately 2.85 m × 3.07 m. The listener's head was positioned against a headrest that contacted only the back of the head while the listener sat in a straight-backed chair. Earphones were calibrated in a 6-cc coupler (Brüel & Kjaer 4152) and it was determined that 1 V at 1 kHz was equal to 116 dB SPL. As the response of the earphones was relatively flat across a wide range of frequencies, no filter was applied when computing the rms level of wide-band signals. Loudspeaker levels were calibrated using a Brüel & Kjaer precision sound level meter Type 2203 with the microphone positioned at the approximate location of a listener's ear. Individual HRTFs were not measured. Therefore, absolute level at each listener's ear in the loudspeaker condition is likely to vary slightly. However, it is not expected that small absolute level differences would have an impact on the binaural-to-monaural loudness ratios because relative levels would remain constant for each listener. Video was displayed on a 17 inch flat-panel monitor placed directly below and in front of the loudspeaker. Listeners typed in magnitude estimates for each stimulus presentation and then the screen was cleared so that no prior estimates were visible to the listener.

C. Listeners

The subjects consisted of 12 normal-hearing listeners (6 males, 6 females), ages 20–33 yrs. All had bilaterally normal thresholds less than 15 dB HL (ANSI, 2004) at audiometric frequencies 250–8000 Hz and no asymmetry greater than 5 dB between the ears. All listeners had normal immittance measures, and histories consistent with normal hearing. Only one had previous experience making loudness judgments. Informed consent was obtained from all listeners.

D. Statistical analysis

In order to test the primary hypothesis and examine additional *post hoc* effects of other factors, a within-subject design was used to examine the following factors: Two presentation modes (video combined with audio, audio only), two transducer types (loudspeaker, earphones), number of ears (monaural, binaural), and five sound levels. A repeated-measures General Linear Model examining the mean logarithms of loudness estimates collapsed across trial using a

four-way analysis of variance (ANOVA) with Huynh–Feldt sphericity correction (SPSS 17.0) was applied to the data to examine whether video combined with audio presented via a loudspeaker results in a smaller binaural-to-monaural loudness ratio than the other conditions. These ratios were examined by looking at interactions including the number of ears as a factor. The loudness estimates were log-transformed to account for the log-normal distribution typically seen in magnitude estimates of loudness. However, binaural-to-monaural loudness ratios presented throughout are the ratios of the actual values of the estimates. (In other words, pairs of judgments that yield magnitude estimates that are twice as large for binaural sounds as otherwise equivalent monaural sounds would result in a binaural-to-monaural ratio of 2.0.)

Because the procedure, magnitude estimation, is expected to yield a logarithmic assessment of loudness, all analyses were performed on the logarithms of the magnitude estimates. In addition to the primary hypothesis, the statistical model was designed to include several factors for *post hoc* analysis and give insight into the procedure and the results in order to examine whether these individual factors, or interactions, resulted in differences in BLS. The data were collapsed across the three trials for each condition, the four words, and the four talkers.

A set of *post hoc*, paired-sample, one-tail *t*-tests using individual listener binaural-to-monaural ratios were conducted to further test the hypothesis that conditions in which speech with video combined with audio presented via a loudspeaker yielded a ratio significantly smaller than all other conditions. The Holm–Bonferroni method (Holm, 1979) was used to control the family-wise error rate for the three comparisons.

III. RESULTS

Figure 1 shows binaural-to-monaural loudness ratios collapsed into four conditions for 12 listeners as a function of level. The four conditions are: Video combined with audio presented via earphones, only audio presented via

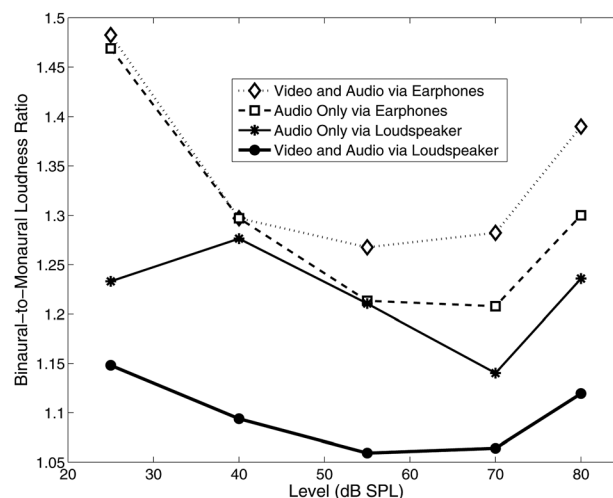


FIG. 1. Ratio of geometric-mean-binaural to geometric-mean-monaural magnitude estimates for 12 listeners as a function of SPL for all conditions shown in the inset.

TABLE I. Overall binaural-to-monaural loudness ratios of mean loudness judgments and standard errors of the mean of the ratios for various listening conditions (see text).

Condition	Mean	
	binaural-to-monaural ratio	Standard error
Overall	1.24	0.05
Earphones	1.32	0.07
Loudspeaker	1.16	0.05
Video combined with audio	1.21	0.05
Audio Only	1.26	0.06
Video combined with audio via earphones	1.34	0.06
Video combined with audio via loudspeaker	1.09	0.04
Audio only via earphones	1.29	0.07
Audio only via loudspeaker	1.22	0.09

earphones, only audio presented via a loudspeaker, and video combined with audio presented via a loudspeaker. Figure 1 and the corresponding table clearly show that the condition in which speech with video combined with audio is presented via a loudspeaker has a smaller binaural-to-monaural loudness ratio at all levels than any of the other conditions. Table I provides an expanded summary of the binaural-to-monaural ratios for a number of conditions collapsed across level that cannot be seen in Fig. 1. These ratios were computed for each condition using the geometric mean of the binaural magnitude estimates for all listeners and the geometric mean of the monaural magnitude estimates for all listeners. The right column shows the standard errors of the mean of the binaural-to-monaural magnitude-estimate ratios. These data indicate that the standard errors for the binaural-to-monaural ratios across individual listeners are relatively small.

Figure 2 shows geometric-mean magnitude estimates as a function of level for all eight conditions. The monaural/binaural pairs are offset arbitrarily vertically and equal SPL points are offset horizontally to increase readability. These results show that all eight functions have similar slopes as a function of level, indicating that loudness grows at an approximately equal rate for all conditions. Standard errors are approximately the same across all conditions.

The observations made from Figs. 1 and 2 are supported by the statistical analysis. In order to test the hypothesis that speech presented under a more ecological condition (i.e., video combined with audio presented via a single loudspeaker) results in less BLS than speech presented via a loudspeaker without visual cues or speech presented via earphones with or without visual cues, the statistical result of interest was the interaction of transducer type (earphones or loudspeaker) \times presentation mode (video combined with audio or audio only) \times number of ears (monaural or binaural). This interaction was significant [$F(1,11) = 15.73$, $p < 0.01$]. Therefore, the statistical model supports the primary hypothesis and the differences observed in Fig. 1 and Table I.

In addition to the primary result, some expected factors were significant: Loudness changed significantly with increasing presentation level [$F(4,44) = 160.102$, $p < 0.01$] and number of ears [$F(1,11) = 25.464$, $p < 0.01$]. (In other words, as sound level increases, loudness increases and binaural sounds are judged louder than their otherwise identical monaural counterparts.) Given the support for the hypothesized difference between the video combined with audio condition presented via a loudspeaker and all other conditions, it was not surprising to find that transducer type \times number of ears was also significant [$F(1,11) = 14.486$, $p < 0.01$]. The interaction of the number of ears \times level was also significant

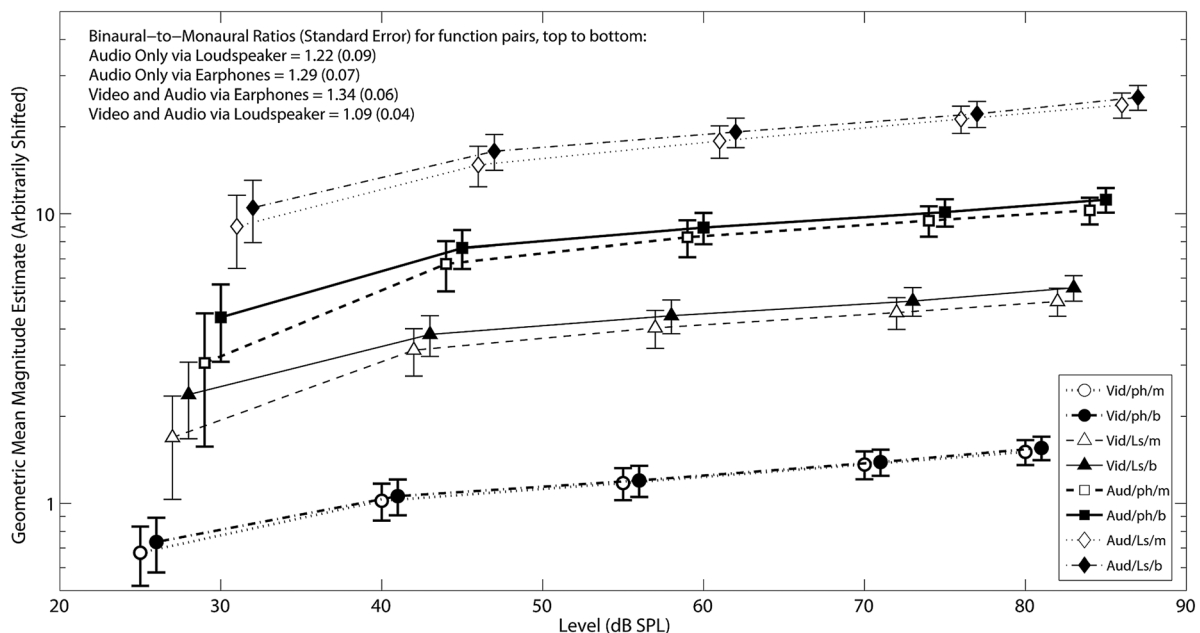


FIG. 2. Geometric-mean magnitude estimates as a function of level for all eight conditions (see bottom-right inset). The vertical bars show the standard errors of the mean for each point. Monaural/binaural function pairs are offset from one another arbitrarily vertically and equal-SPL points are offset horizontally to increase readability. Despite this offset, all data points represent measurements at 25, 40, 55, 70, or 80 dB SPL. Average binaural-to-monaural ratios and the standard errors (shown in parentheses) for binaural-to-monaural ratios across listeners are shown in the top left for the four pairs of functions.

TABLE II. One-tailed, *post hoc*, paired t-test comparisons between ratios for 12 listeners.

Condition 1	Condition 2	<i>t</i>	<i>p</i>	Holm–Bonferroni α
Video combined with audio via loudspeaker	Video combined with audio via earphones	3.531	0.003	0.017
Video combined with audio via loudspeaker	Audio only via earphones	2.961	0.007	0.025
Video combined with audio via loudspeaker	Audio only via loudspeaker	1.896	0.04	0.05

[$F(4,4) = 9.974$, $p < 0.01$]. This indicates that the binaural-to-monaural ratio was not constant as a function of level for the speech stimuli.

Table II shows the results of the *post hoc* *t*-tests for all pairings of conditions. A Holm–Bonferroni correction was used. The condition in which the video combined with audio was presented via a loudspeaker was found to be significantly different than all three other conditions. These results support the present hypothesis that the most ecologically valid condition (i.e., speech presented with video combined with audio presented via a loudspeaker) results in a lower binaural-to-monaural loudness ratio than the other three conditions.

IV. DISCUSSION

A. Comparison with data in the literature

The present results lend support to the proposed binaural loudness constancy hypothesis (i.e., that speech presented under conditions that are closer to typical daily environments result in less BLS than speech presented under less ecologically valid conditions). Specifically, the amount of BLS obtained at the same SPL is significantly smaller for (1) speech from a visually present talker presented via a loudspeaker than any of the other three conditions, and (2) loudspeaker presentation than for earphone presentation. These results corroborate the findings of Epstein and Florentine (2009). The reader may question the strength of this corroboration because the present authors are the same as those in the previous study. The authors, however, were the only identical factor between the two studies; the stimuli, procedure, experimental apparatus, subjects, research assistants, and testing sites (research laboratory vs clinical setting) were different. Therefore, agreement between the results of the two studies, while expected, was not guaranteed.

The present data are also in quantitative agreement with the data of Epstein and Florentine (2009). Most notably, the binaural-to-monaural loudness ratio for the video combined with audio presentation of speech via a loudspeaker in the present experiment is 1.09 and the mean loudness ratio for a similar condition in Epstein and Florentine’s (2009) experiment was 1.08. The binaural-to-monaural loudness ratio across both conditions (video combined with audio, audio) for earphone presentation in the present experiment is 1.32, which is within the range (i.e., 1.29–1.7) of most other studies referenced in the introduction (e.g., Epstein and Florentine, 2009; Marozeau *et al.*, 2006; Scharf and Fishken, 1970; Zwicker and Zwicker, 1991).

There are some similarities and differences between the present data and those obtained by Cox and Gray (2001). One similarity is that their data showed less difference between the

binaural and monaural loudness functions for their loudspeaker presentation than for their earphone presentation. Their loudspeaker presentation had a small, but significant, average difference of 1.06 dB between the two loudness functions, whereas the earphone presentation had an average significant difference of 3.75 dB. In both studies, the amount of BLS obtained at the same SPL is smaller for the loudspeaker presentation than for the earphone presentation. Although Cox and Gray’s data indicate that BLS is less for speech presented via a loudspeaker than for speech presented via earphones, both of their conditions were presented without visual cues. In the present experiment, loudspeaker presentation alone without visual cues *did not* significantly reduce BLS when compared with earphone presentation without visual cues. Two important differences between the studies could be responsible for the different outcomes: Stimuli and methods of measurement. One notable difference is that Cox and Gray used speech stimuli with 5-s durations that may have served to create a more ecologically valid listening situation than the short spondee stimuli used in the present experiment. A longer time to aurally localize the sound source may have had a similar effect that visual cues provide for short speech sounds. It is also important not to overemphasize the differences between the two studies. Cox and Gray’s experiment was designed to answer a clinical question and they used categorical loudness comfort judgments. They measured loudness comfort, which is a different attribute than pure loudness. The present experiment was specifically designed to measure loudness. Although this may seem like a small difference, there is a long history showing that differences in methods of measurements influence loudness results (Marks and Florentine, 2011).

The speech data from the present experiment are inconsistent with the binaural equal-loudness-ratio hypothesis (BELRH, Marozeau *et al.*, 2006; Marozeau and Florentine, 2009), which states that the loudness ratio between equal SPL binaural and monaural tones is independent of SPL. The analysis for the speech data indicates that the binaural-to-monaural ratio is not constant as a function of level. The ratios were: 1.32 at 25 dB SPL, 1.24 at 40 dB SPL, 1.19 at 55 dB SPL, 1.17 at 70 dB SPL, and 1.26 at 80 dB SPL. This indicates that there may be something different about the way tones and speech are processed.

Several studies have reported differences in the slope of loudness functions for tones and speech (e.g., Mendel *et al.*, 1969; Pollack, 1952). The determining factors for this difference are likely to be acoustic and non-acoustic, as they are for loudness constancy with distance from a sound source (for review, see Zahorik *et al.*, 2005). In room environments, sound level and direct-to-reverberant energy ratio are the primary acoustic cues for distance, although they appear to be weighted

differently for speech and noise signals. One possible reason for the shallow loudness functions in the present study is that the speech stimuli were recorded at a normal vocal effort level and played back at five different SPLs. Speech stimuli with steady vocal effort yield shallower loudness functions than pure tones, noise, and speech with varied vocal effort (Mendel *et al.*, 1969). This is not surprising because listeners can easily judge vocal effort from the spectrum of speech. For example, a recorded whisper will sound like a whisper even if played back at a high level and perceived as loud.

B. Multisensory interactions

The finding of almost no BLS in the most ecologically valid condition—and therefore the confirmation of the existence of binaural loudness constancy—should not be surprising when viewed in the wider context of multisensory interactions. The McGurk effect (McGurk and McDonald, 1976) is a well-known example of how perception can involve more than one sense; if a subject observes a talker articulating /ga/ and simultaneously hears /ba/, the subject reports perceiving /da/. Visual capture (aka ventriloquism) is a well-known example of visual location taking precedence over the spatial location of an auditory sound source (i.e., the talking dummy is perceived, see Choe *et al.*, 1975). Research on the influence of one modality on another abounds (e.g., vision on hearing; for a review of multisensory interactions in ratings of loudness see Fastl and Florentine, 2011). Sivonen and Ellermeier (2011) point out that binaural loudness constancy may be analogous to the visual phenomenon of binocular brightness judgments in which closing one eye does not make the world appear markedly less bright.

One potential confound in the present study was the influence of brightness on loudness. Odgaard *et al.* (2004) showed that white noise presented with light tends to be rated as louder than noise presented alone. Because listeners in the present experiment were positioned in front of a lighted 17 inch flat-panel monitor for all the measurements, the influence of brightness would have been about constant for all measurements. Although it is possible that the listeners could have closed their eyes to make the audio-only judgments and then opened their eyes to make their responses, there is no evidence of sensory facilitation in the data.

C. Possible cues for binaural loudness constancy

How do stimulus conditions differ between ecologically valid environments and laboratories? One important difference is the presence of reverberation. Reverberant energy distorts the acoustic signal as it travels from its source to the listener. Reverberation can alter the spatial cues, such as interaural time and level differences, and interfere with auditory object formation (Darwin and Hukin, 2000; Lavandier and Culling, 2008; Nabelek *et al.*, 1989). Mohrmann (1939) observed that loudness of conversational speech changes little with distance from a sound source in reverberant rooms. One possible explanation is that auditory processing could remove reverberation prior to the stage at which loudness judgments are made (Stecker and Hafter, 2000).

The phenomenon by which loudness remains constant as distance from the sound source is changed is known as loudness constancy (see Zahorik and Wightman, 2001) because of its similarity to perceptual constancy (e.g., the perceived size of an object does not change when viewed at a distance; for a basic review of perceptual constancy see Goldstein, 2009). Zahorik and Wightman's (2001) experiment indicates that loudness constancy for distance is related to reverberant sound energy, which remains relatively constant with distance from a sound source in normally reverberant rooms. Consistent with this contention is the finding that loudness constancy is absent under anechoic conditions. In anechoic conditions, SPL is likely to be the primary cue; there is an orderly relation between distance and loudness (Stevens and Guirao, 1962; Petersen, 1990). It is likely that reverberation is a necessary cue for binaural loudness constancy, but not a sufficient one.

One intriguing study showed that even when a true sound source remains stationary and sound is presented at a fixed level, there is evidence that perceived location affects loudness (Mershon *et al.*, 1981). In the present experiment, binaural loudness constancy was only observed in the condition in which visual and auditory cues were presented simultaneously. Therefore, a prerequisite condition for binaural loudness constancy appears to be a visually present talker paired with the auditory stimulus presented in a typical room. Either condition alone—just the visually present talker heard via earphones or the visually absent talker heard via a loudspeaker—did not result in binaural loudness constancy.

D. Implications for loudness models

The present data provide a challenge for current loudness models. The concept of binaural inhibition has been used as a theoretical construct by different investigators to explain aspects of BLS (e.g., Gigerenzer and Strube, 1983; Hirsh, 1948; Moore and Glasberg, 2007; and most recently Glasberg and Moore, 2010). Binaural loudness inhibition assumes that a signal in one ear reduces the loudness evoked by a signal in the other ear. Although binaural inhibition may exist, it cannot in its current formulation explain binaural loudness constancy.

Although progress has been made in modeling diotic and dichotic loudness for normal-hearing listeners (see Glasberg and Moore, 2010), current loudness models for steady-state or time-varying sounds cannot account for binaural loudness constancy (Glasberg and Moore, 2002; Glasberg and Moore, 2010; Chalupper and Fastl, 2002). Current models generally use only the present physical characteristics of sound, ignoring context and the influence of other modalities. The present data indicate that these overlooked ecologically relevant factors, such as visual cues resulting in binaural loudness constancy, can have a significant impact on perception. It is also insufficient to ignore the context and content carried by a stimulus. There are complex influences of context on loudness in daily environments (for a review, see Fastl and Florentine, 2011) and it is likely that the perception of loudness occurs subsequent to perceptual organization (see McAdams *et al.*, 1998). Furthermore, present findings indicate that caution

should be used when applying loudness standards (ANSI, 2007; ISO, 1975) to daily environmental conditions.

E. Clinical relevance

The influence of BLS is a common concern in binaural hearing-aid fittings. According to the survey of Kochkin *et al.* (2010) of over 3000 hearing-aid users, only 67% of hearing-aid users are satisfied with their comfort with loud sounds. It is clear that loudness functions for listeners with hearing losses show a considerable individual difference (for a review, see Marozeau and Florentine, 2007) and differences in loudness-comfort growth functions for speech and tones (Cox *et al.*, 1997). Although there are many factors that influence loudness in hearing losses (for a review, see Smeds and Leijon, 2011), BLS may be an important one that appears to differ among individuals (Marozeau and Florentine, 2009; Whilby *et al.*, 2006). There is currently no method to ascertain if hearing-aid users experience BLS in a manner similar to individuals with normal hearing.

There is much that we do not know about BLS in hearing losses. Is listening with hearing aids more similar to listening with earphones, loudspeaker, or a combination of both? Do hearing-aid users adapt over time? Is it possible to adapt if advanced processing schemes in hearing aids keep changing aspects of the auditory signal that are used as cues by listeners? Indeed, there are many unanswered questions and there are highly likely to be individual differences among hearing-aid users. With our current knowledge of BLS, it is impossible to predict optimal loudness for any individual hearing-aid user. The present data lend further support to Cox and Gray's (2001) recommendation to verify loudness perception with loudspeakers after hearing-aid fittings. The present study also supports the concept of performing these fittings with accompanying visual stimuli.

V. CONCLUSIONS

Although loudness is a one-dimensional concept in theory and research, it is a multi-dimensional concept as it is experienced in daily environments (Florentine, 2011). Taken together, the available data support the concept of binaural loudness constancy in which speech (perhaps all natural sounds) from a visually present source heard under ecologically valid conditions is not much louder when listening with two ears than when listening with only one ear (see Epstein and Florentine 2009). Although the present experimental design did not permit an examination of all variables, which need to be assessed in detail in future studies, the present study sheds light on binaural loudness constancy. Note that this phenomenon is not only due to the visual cues of a person articulating spondees; the data from video combined with audio presentation via earphones did not result in a significantly smaller loudness ratio than the same stimuli presented without video. It is also not only due to the transducer; data from the audio only presentation via loudspeaker did not result in a significantly smaller ratio than the same stimuli presented via earphones. In contrast, however, Cox and Gray's (2001) categorical loudness-comfort judgments did indicate that the presentation of speech stimuli with 5-s durations resulted in reduced BLS, even without a vis-

ually present talker. The contributions of the stimulus type, visual presence of talker, methods of measurement, and their interactions merit further examination.

In daily environments, there are a multitude of cues that a listener may use. In survival situations, it could be advantageous if a listener were able to accurately judge the distance of an environmentally relevant sound whether listening with one or two ears. The results of the present experiment are consistent with the importance of ecological validity in loudness research, which could change how perception of loudness is understood.

ACKNOWLEDGMENTS

Kevin Reilly provided assistance and equipment for stimulus recordings. Tao Cui, Eliza Floyd, Abigail Seaser, Yui Anzai, and Maria Zilberberg assisted with data acquisition. Tao Cui also assisted with stimulus preparation and Yui Anzai also assisted with data analysis. Whitney Danse, Amy Levasseur, Thomas DiCicco, and Michael Epstein served as talkers. Julia Buus Florentine and Andre Lira provided helpful comments, as did two anonymous reviewers.

- 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.
- ANSI. (2004). "American National Standard Specification for Audiometers" (American National Standards Institute, New York).
- ANSI. (2007). "American National Standard Procedure for the Computation of Loudness of Steady Sounds" (American National Standards Institute, New York).
- Brungart, D. S., and Scott, K. R. (2001). "The effects of production and presentation level on the auditory distance perception of speech," *J. Acoust. Soc. Am.* **110**, 425–440.
- Chalupper, J., and Fastl, H. (2002). "Dynamic Loudness Model (DLM) for normal and hearing-impaired listeners," *Acta. Acust. Acust.* **88**, 378–386.
- Choe, C. S., Welsh, R. B., Gilford, R. M., and Juola, J. F. (1975). "The ventriloquist effect: Visual dominance or response bias?," *Percept. Psychophys.* **18**(1), 55–60.
- Cox, R. M., Alexander, G. C., Taylor, I. M., and Gray, G. A. (1997). "The contour test of loudness perception," *Ear Hear.* **18**, 388–400.
- Cox, R. M., and Gray, G. A. (2001). "Verifying loudness perception after hearing aid fitting," *Am. J. Audiol.* **10**, 91–98.
- Darwin, C. J., and Hukin, R. W. (2000). "Effects of reverberation on spatial, prosodic, and vocal-tract size cues to selective attention," *J. Acoust. Soc. Am.* **108**, 335–342.
- Edmonds, B. A., and Culling, J. F. (2009). "Interaural correlation and the binaural summation of loudness," *J. Acoust. Soc. Am.* **125**, 3865–3870.
- Epstein, M., and Florentine, M. (2009). "Binaural loudness summation for speech and tones presented via earphones and loudspeakers," *Ear Hear.* **30**, 234–237.
- Fastl, H., and Florentine, M. (2011). "Loudness in daily environments," in *Loudness*, edited by M. Florentine, A. N. Popper, and R. R. Fay (Springer, New York), pp. 199–222.
- Fletcher, H., and Munson, W. A. (1933). "Loudness, its definition, measurement and calculation," *J. Acoust. Soc. Am.* **5**, 82–108.
- Florentine, M. (2011). "Loudness," in *Loudness*, edited by M. Florentine, A. N. Popper, and R. R. Fay (Springer, New York), pp. 1–16.
- Gigerenzer, G., and Strube, G. (1983). "Are there limits to binaural additivity of loudness?," *J. Exp. Psychol. Hum. Percept. Perform.* **9**, 126–136.
- Glasberg, B. R., and Moore, B. C. (2002). "A model of loudness applicable to time-varying sounds," *J. Audio Eng. Soc.* **50**, 331–342.
- Glasberg, B. R., and Moore, B. C. (2010). "The loudness of sounds whose spectra differ at the two ears," *J. Acoust. Soc. Am.* **127**, 2433–2440.
- Goldstein, E. B. (2009). *Sensation and Perception* (Wadsworth, Pacific Grove, CA), pp. 1–459.
- Hellman, R. P. (1991). "Loudness scaling by magnitude scaling: Implications for intensity coding," in *Ratio Scaling of Psychological Magnitude*:

- In Honor of the Memory of S. S. Stevens*, edited by G. A. Gescheider and S. J. Bolanowski (Erlbaum, Hillsdale, NJ), pp. 215–228.
- Hirsh, I. J. (1948). “Binaural summation and interaural inhibition as a function of the level of the masking noise,” *J. Acoust. Soc. Am.* **56**, 205–213.
- Holm, S. (1979). “A simple sequentially rejective multiple test procedure,” *Scand. J. Stat.* **6**(2), 65–70.
- ISO. (1975). “Acoustics—Method for calculating loudness level” (International Organization for Standardization, Geneva).
- Kochkin, S., Beck, D., Christensen, L., Compton-Conley, C., Kricos, P., Fligor, B., McSpaden, J., Mueller, H., Nilsson, M., Northern, J., Powers, T., Sweetow, R., Taylor, B., and Turner, R. (2010). “MarkeTrak VIII: The impact of the hearing healthcare professional on hearing aid user success,” *Hear. Rev.* **17**, 12–34.
- Lavandier, M., and Culling, J. F. (2008). “Speech segregation in rooms: monaural, binaural, and interacting effects of reverberation on target and interferer,” *J. Acoust. Soc. Am.* **123**, 2237–2248.
- Marks, L. E. (1978). “Binaural summation of the loudness of pure tones,” *J. Acoust. Soc. Am.* **64**, 107–113.
- Marks, L. E., and Florentine, M. (2011). “Measurement of loudness, Part I. Methods, problems, and pitfalls,” in *Loudness*, edited by M. Florentine, A. N. Popper, and R. R. Fay (Springer, New York), pp. 57–88.
- Marozeau, J., Epstein, M., Florentine, M., and Daley, B. (2006). “A test of the binaural equal-loudness-ratio hypothesis for tones,” *J. Acoust. Soc. Am.* **120**, 3870–3877.
- Marozeau, J., and Florentine, M. (2007). “Loudness growth in individual listeners with hearing losses: A review,” *J. Acoust. Soc. Am.* **122**, EL81–EL87.
- Marozeau, J., and Florentine, M. (2009). “Testing the binaural equal-loudness-ratio hypothesis with hearing-impaired listeners,” *J. Acoust. Soc. Am.* **126**, 310–317.
- McAdams, S., Botte, M. C., and Drake, C. (1998). “Auditory continuity and loudness computation,” *J. Acoust. Soc. Am.* **103**, 1580–1591.
- McGurk, H., and McDonald, J. (1976). “Hearing lips and seeing voices,” *Nature* **264**, 746–748.
- Mendel, M. I., Sussman, H. M., Merson, R. M., Naeser, M. A., and Minifie, F. D. (1969). “Loudness judgments of speech and nonspeech stimuli,” *J. Acoust. Soc. Am.* **46**, 1556–1561.
- Mershon, D. H., Desaulniers, D. H., Kiefer, S. A., Amerson, T. L., Jr., and Mills, J. T. (1981). “Perceived loudness and visually-determined auditory distance,” *Perception* **10**, 531–543.
- Mohrmann, K. (1939). “Lautheitskonstanz im entfernungswechsel (Loudness constancy with changing distance),” *Z. Psychol.* **145**, 145–199.
- Moore, B. C., and Glasberg, B. R. (2007). “Modeling binaural loudness,” *J. Acoust. Soc. Am.* **121**, 1604–1612.
- Nabelek, A. K., Letowski, T. R., and Tucker, F. M. (1989). “Reverberant overlap- and self-masking in consonant identification,” *J. Acoust. Soc. Am.* **86**, 1259–1265.
- Odgaard, E. C., Arie, Y., and Marks, L. E. (2004). “Brighter noise: sensory enhancement of perceived loudness by concurrent visual stimulation,” *Cogn. Affect. Behav. Neurosci.* **4**, 127–132.
- Petersen, J. (1990). “Estimation of loudness and apparent distance of pure tones in a free-field,” *Acustica* **70**, 61–65.
- Pollack, I. (1952). “The loudness of bands of noise,” *J. Acoust. Soc. Am.* **24**, 533–538.
- Scharf, B. (1969). “Dichotic summation of loudness,” *J. Acoust. Soc. Am.* **45**, 1193–1205.
- Scharf, B., and Fishken, D. (1970). “Binaural summation of loudness: reconsidered,” *J. Exp. Psychol.* **86**, 374–379.
- Sivonen, V. N. P., and Ellermeier, W. (2008). “Binaural loudness for artificial-head measurements in directional sound fields,” *J. Audio Eng. Soc.* **56**, 452–461.
- Sivonen, V. P., and Ellermeier, W. (2011). “Binaural loudness,” in *Loudness*, edited by M. Florentine, A. N. Popper, and R. R. Fay (Springer, New York), pp. 169–198.
- Smeds, K., and Leijon, A. (2011). “Loudness and hearing loss,” in *Loudness*, edited by M. Florentine, A. N. Popper, and R. R. Fay (Springer, New York), pp. 223–260.
- Stecker, G. C., and Hafter, E. R. (2000). “An effect of temporal asymmetry on loudness,” *J. Acoust. Soc. Am.* **107**, 3358–3368.
- Stevens, S. S., and Guirao, M. (1962). “Loudness, reciprocity, and partition scales,” *J. Acoust. Soc. Am.* **34**, 1466–1471.
- Warren, R. M. (1973). “Anomalous loudness function for speech,” *J. Acoust. Soc. Am.* **54**, 390–396.
- 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.
- Zahorik, P., Brungart, D. S., and Bronkhorst, A. W. (2005). “Auditory distance perception in humans: A summary of past and present research,” *Acta Acust. Acust.* **91**, 409–420.
- Zahorik, P., and Wightman, F. L. (2001). “Loudness constancy with varying sound source distance,” *Nat. Neurosci.* **4**, 78–83.
- 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.