Binaural Loudness Constancy

John F. Culling and Helen Dare

Abstract In binaural loudness summation, diotic presentation of a sound usually produces greater loudness than monaural presentation. However, experiments using loudspeaker presentation with and without earplugs find that magnitude estimates of loudness are little altered by the earplug, suggesting a form of loudness constancy. We explored the significance of controlling stimulation of the second ear using meatal occlusion as opposed to the deactivation of one earphone. We measured the point of subjective loudness equality (PSLE) for monaural vs. binaural presentation using an adaptive technique for both speech and noise. These stimuli were presented in a reverberant room over a loudspeaker to the right of the listener, or over lightweight headphones. Using the headphones, stimuli were either presented dry, or matched to those of the loudspeaker by convolution with impulse responses measured from the loudspeaker to the listener position, using an acoustic manikin. The headphone response was also compensated. Using the loudspeaker, monaural presentation was achieved by instructing the listener to block the left ear with a finger. Near perfect binaural loudness constancy was observed using loudspeaker presentation, while there was a summation effect of 3–6 dB for both headphone conditions. However, only partial constancy was observed when meatal occlusion was simulated. These results suggest that there may be contributions to binaural loudness constancy from residual low frequencies at the occluded ear as well as a cognitive element, which is activated by the knowledge that one ear is occluded.

Keywords Loudness summation · Perceptual constancy · Meatal occlusion · Monaural · Virtual acoustics

1 Introduction

When the same sound is presented to both ears, it is perceived to be louder than when it is presented to one ear only (Fletcher and Munson 1933). As a result, the loudness of diotically and monaurally presented stimuli, usually tones or noise, has

J. F. Culling (✉) · H. Dare
School of Psychology, Cardiff University, Tower Building, Park Place, Cardiff, CF10 3AT, UK
e-mail: CullingJ@cf.ac.uk

been found to be equivalent when the monaural stimulus is between 3 and 10 dB more intense (Fletcher and Munson 1933; Reynolds and Stevens 1960; Zwicker and Zwicker 1991; Sivonen and Ellermeier 2006; Whilby et al. 2006; Edmonds and Culling 2009). Models of loudness therefore incorporate this phenomenon in various ways (Zwicker and Scharf 1965; Moore and Glasberg 1996a, b; Moore and Glasberg 2007). These models can be used in hearing-aid fitting (e.g. Moore and Glasberg 2004), so hearing aids fitted using these models will have relatively reduced gain when fitted bilaterally in order to achieve a comfortable maximum loudness level. However, some recent studies have cast doubt on the ecological validity of the loudness summation literature.

Cox and Gray (2001) collected loudness ratings (7 categories from "very soft" to "uncomfortable") for speech at different sound levels when listening monaurally and binaurally. These two modes of listening were compared by using either one or two earphones and, using a loudspeaker, by occluding one ear with a plug and ear muff. They found that results with one or two earphones produced a conventional loudness summation effect, whereby the mean rating was substantially higher at each sound level for binaural presentation. However, when listening to an external source in the environment (the loudspeaker) there was much less summation effect: occlusion of one ear had little effect on the loudness ratings compared to listening binaurally. This experiment showed for the first time that the loudness of external sounds may display constancy across monaural and binaural listening modes. However, the methods used were clinically oriented and difficult to compare with conventional psychophysical measurements.

Epstein and Florentine (2009, 2012) conducted similar tests, but using standard loudness estimation procedures and speech (spondees) either with or without accompanying video of the speaker's face. They also observed loudness constancy, but only when using the audiovisual presentation. Their tentative conclusion was that loudness constancy may only occur using stimuli of relatively high ecological validity. Ecological validity may be enhanced when an external source is used, when that source is speech, particularly connected speech rather than isolated words, and when accompanied by coherent visual cues. Since all these conditions are fulfilled when listening to someone in real life, the phenomenon of loudness constancy can be compellingly, if informally, demonstrated by simply occluding one ear with a finger when listening to someone talk; most people report that no change in loudness is apparent. The present study was inspired by this simple demonstration technique.

A limitation of previous studies is that the use of ear plugs and muffs means that different listening modes cannot be directly compared. It is commonplace in psychophysical testing to match the loudness of different stimuli by listening to them in alternation, but it is impractical to insert/remove an ear plug between one presentation interval and another. As a result comparisons are made over quite long time intervals. In contrast, a finger can be applied to the meatus in less than a second, so it is possible to perform a 2-interval, forced-choice procedure provided that the monaural interval is always the same one so that the listener can learn the routine of blocking one ear for the same interval of each trial. The present experiment

adopted this technique and also explored the idea that ecological validity plays a key role by using either speech or noise as a sound source and by creating a close physical match between stimuli presented from a loudspeaker and those presented using headphones.

2 Methods

2.1 Stimuli

There were six conditions, comprised of two stimulus types (speech/noise) and three presentation techniques. The speech stimuli were IEEE sentences (Rothauser et al. 1969) and the noise stimuli were unmodulated noises filtered to have the same long-term excitation pattern (Moore and Glasberg 1987) as the speech. These two stimulus types were presented (1) dry over headphones, (2) from a loudspeaker to the listeners' right or (3) virtually through lightweight open-backed headphones (Sennheiser HD414). Monaural presentation was achieved in (1) and (3) by deactivating the left earphone for the second interval of each trial and in (2) by asking the listener to block their left ear with a finger for the second interval of each trial. For virtual presentation, the stimuli were convolved with binaural room impulse responses (BRIRs) recorded from a manikin (B&K HATS 4100) sitting in the listeners' position and wearing the lightweight headphones with the loudspeaker 1 m to the right in a standard office room with minimal furnishing (reverberation time, RT60 = 650 ms). Impulse responses were also recorded from the headphones and used to derive inverse filters to compensate for the headphone-to-microphone frequency response. Sound levels for the different stimuli were equalised at the right ear of the manikin and, for the reference stimuli, were equivalent to 57 dB(A) as measured at the head position with a hand-held sound-level meter.

2.2 Procedure

Twelve undergraduate-students with no known hearing impairments took part in a single 1 h session. Initially, five practice trials were presented, for which the listeners were simply required to learn the routine of blocking their left ear during the one-second inter-stimulus interval between two bursts of speech-shaped noise presented from a loudspeaker. During the experiment, the ordering of six conditions was randomly selected for each participant.

For each condition, listeners completed two loudness discrimination threshold tasks. These adaptive thresholds served to bracket the point of subjective loudness equality (PSLE): one was the 71 % threshold for identifying when a monaural stimulus was louder than a binaural one, and the other was the 71 % threshold for identifying when the monaural stimulus was quieter than a binaural one. For a given

Fig. 1 The mean summation effect in Expt. 1 in each of the three presentation conditions and for speech and for noise sources. Error bars are one standard error
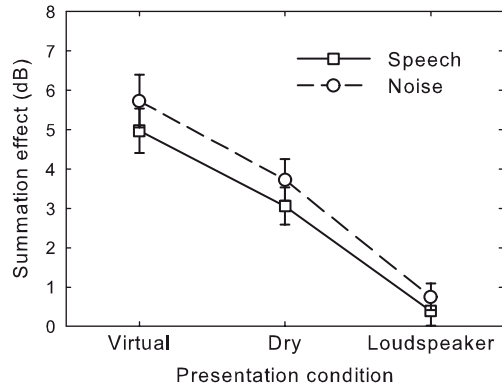
condition, the two adaptive tracks started with the adapted (monaural) sound level 15 dB above and below the level of the reference (binaural) stimulus, respectively. In each task, the listeners were required to identify the louder of two stimuli in a 2-down, 1-up adaptive procedure with six reversals. The two thresholds were averaged to yield a PSLE for that condition.

Within each trial the first interval was always the reference stimulus, while the adapted stimulus was presented in the second interval. The adapted stimulus was thus always the monaural one.

3 Results

Figure 1 shows that there was a strong effect of the presentation condition ($F(2,22)=22.1$, $p<0.001$). The summation effect is less than 1 dB in the Loudspeaker condition, in which monaural presentation was achieved by occluding the meatus with one finger, but is 3–6 dB in the Virtual and Dry conditions, in which monaural presentation was achieved by deactivating one earphone. Summation was also significantly smaller using speech than using noise as a source signal ($F(1,11)=5.6$, $p<0.05$).

4 Discussion

The lack of summation in the Loudspeaker condition supports the observations made by Cox and Gray (2001) and by Epstein and Florentine (2009, 2012), that, when monaural presentation is achieved by occluding one ear in a sound field, binaural summation can be dramatically reduced. The present results demonstrate this phenomenon for the first time using direct comparisons between successive stimuli. Consistent with suggestions that ecological validity is important to the effect, it was a little larger for speech than for noise, but this effect was rather small (0.8 dB).

In order to match the headphone and loudspeaker conditions physically, virtual presentation was used in a reverberant room. This approach resulted in a compelling impression of sound source externalisation for the headphone condition, which may have been accentuated by the presence of a loudspeaker at the appropriate position. A post-experiment test showed that listeners perceived the virtual binaural stimuli as emanating from the loudspeaker rather than the headphones. A consequence of this procedure was that when the left headphone was deactivated for monaural presentation, the externalisation of the sound collapsed, so that listeners found themselves comparing one sound that appeared to be from the loudspeaker with another that, although still reverberant and from the same side, appeared to come from the headphones. It is unclear whether this effect may have caused the apparent enhancement of the loudness summation effect in the Virtual condition.

In each of the studies in which binaural summation has been largely abolished, the second ear has been occluded by some means. The lack of summation could have two logical causes. First, there may be some residual sound energy at the occluded ear that is sufficient to maintain its contribution to binaural loudness. Second, there may be some cognitive element which generates loudness constancy through the knowledge that one ear is occluded. It is important, therefore to determine the amount of residual sound and the role it might play. Cox and Gray used a combination of ear plug and muff and reported that this combination produced a threshold shift of 23 dB at 250 Hz and 52 dB at 4 kHz, indicating that the residual sound was substantially attenuated. Epstein and Florentine used only an ear plug and reported the resulting attenuation caused a threshold shift of 20–24 dB at 1 kHz. In the present study, a finger was used. In order to determine the likely degree of attenuation achieved, we recorded binaural room impulse responses BRIRs with and without an experimenter's finger over the meatus. This was not possible with the B&K HATS 4100 because it has the microphone at the meatus, so KEMAR was used. Several repeated measurements found a consistent attenuation of ~30 dB across most of the frequency spectrum, but progressively declining at low frequencies to negligible levels at 200 Hz and below. Assuming this measurement gives a reasonably accurate reflection of the change in cochlear excitation when a human listener blocks the meatus, the question arises whether this residual sound might be sufficient to support summation.

First, we conducted an analysis of the virtual stimuli by applying the Moore and Glasberg (1996a, b) loudness model. Since similar effects were observed with noise as with speech, it was thought unnecessay to employ the model for time-varying sounds (Glasberg and Moore 2002). Despite presentation of stimuli from one side, the difference in predicted loudness at each unoccluded ear differed by only 4%. The interaural level difference was small due to the room reverberation. When the left ear was occluded, the predicted loudness at this ear fell by 52%. Older models assume that binaural loudness is the sum of that for each ear, but more recent models predict less binaural summation by invoking an inhibitory mechanism (Moore and Glasberg 2007). In either case, attenuation caused by meatal occlusion leading to a 52% drop at one ear seems more than adequate to observe a strong summation effect.

While modelling can be persuasive, we wanted to be sure that the relatively unat-tenuated low frequencies were not responsible for a disproportionate contribution to binaural loudness. A second experiment examined this possibility by simulating occlusion.

5 Methods

Eight listeners took part in a similar experiment, but using a new presentation con-dition that simulated the use of a finger to block the meatus. One of the BRIRs collected with a finger over the ear of a KEMAR was convolved with the source signals to form the monaural stimulus of a Virtual Finger condition. This condi-tion was contrasted with a replication of the previous Virtual condition in which monaural presentation was achieved by silencing the left ear. This condition is now labelled Silence, so that the condition labels reflect the method of attenuating the left ear. For consistency, all BRIRs for the second experiment were collected using KEMAR. The binaural stimulus for each condition was identical. The same speech and noise sources were used.

6 Results

Figure 2 shows the mean summation effect from each of the four conditions of the second experiment. The Silence condition produces a comparable summation effect to the similarly constructed Virtual condition of Expt. 1. However, while the Virtual Finger condition does not result in an abolition of this summation effect, it does re-sult in a substantial reduction in its size ($F(1,7)=102, p<0.001$). In contrast to Expt. 1, the noise stimulus produces a smaller summation effect than the speech stimulus ($F(1,7)=22, p<0.005$). The interaction fell short of significance.

Fig. 2 The mean summa-tion effect in Experiment 2 in each of the presentation conditions and for speech and for noise sources. Error bars are one standard error
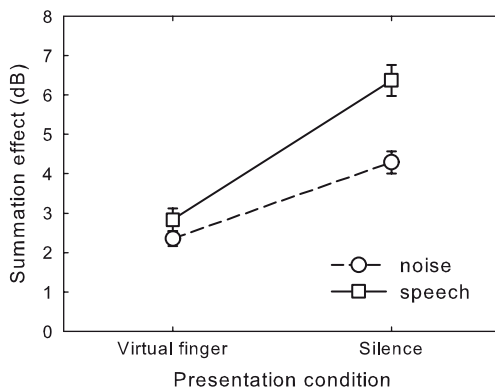
7   Discussion

Since the Silence and Virtual-Finger conditions of experiment 2 produced significantly different summation effects, it seems that a portion of the loudness constancy effect observed in experiment 1 may have been mediated by the residual sound at the occluded ear. Both the Silence and the Virtual-Finger conditions produced a collapse in externalisation for the monaural case, so it seems unlikely that the difference in effect can be attributed to this cause.

On the other hand, 2.5–3 dB of summation was still observed in the Virtual-Finger condition, whereas less that 1 dB was observed in the Loudspeaker condition of experiment 1. It appears, therefore, that the listeners' awareness of meatal occlusion, which would only be present in the first case, may still play a role. A major caveat to these conclusions is that KEMAR was not designed to realistically simulate the effects of bone-conducted sound, which likely plays a major role in listening in a sound field with an occluded ear. While probe microphones in the ear canal might more accurately record the stimulation at the tympanic membrane, they would still not capture the stimulation occurring at the cochlea.

The small effect of stimulus type observed in experiment 1, was reversed in experiment 2. Taking these small and inconsistent effects together, we found no evidence that binaural loudness constancy is greater for more ecologically valid stimuli such as connected speech. Indeed, for both speech and noise, occlusion of the ear led to little reduction in loudness (a small summation effect), suggesting that loudness constancy occurs independently of stimulus type. It should be noted, however, that quite prominent reverberation was present in our experiment, which reliably cued the existence of an external sound source.

8   Conclusions

Experiment 1 demonstrated for the first time that binaural loudness constancy can be observed in a loudness matching task using direct comparisons of loudness between an occluded and unoccluded second ear. Experiment 2 showed that this effect could be partially mediated by the residual sound at the occluded ear, but the remaining effect would seem attributable to the listeners' awareness of the occlusion.

References

Cox RM, Gray GA (2001) Verifying loudness perception after hearing aid fitting. Am J Audiol 10:91–98

Edmonds BA, Culling JF (2009) Interaural correlation and the binaural summation of loudness. J Acoust Soc Am 125:3865–3870

Epstein M, Florentine M (2009) Binaural loudness summation for speech and tones presented via earphones and loudspeakers. Ear Hear 30:234–237

Epstein M, Florentine M (2012) Binaural loudness summation for speech and tones presented via earphones and loudspeakers. J Acoust Soc Am 131:3981–3988

Fletcher H, Munson WA (1933) Loudness, its definition, measurement and calculation. J Acoust Soc Am 5:82–108

Glasberg BR, Moore BCJ (2002) Model of loudness applicable to time-varying sounds. J. Audio Eng Society 53:331–342

Moore BCJ, Glasberg BR (1987) Formulae describing frequency selectivity as a function of frequency and level, and their use in calculating excitation patterns. Hear Res 28:209–225

Moore BCJ, Glasberg BR (1996a) A revision of Zwicker's loudness model. Acustica 82:335–345

Moore BCJ, Glasberg BR (1996b) A model for the prediction of thresholds, loudness, and partial loudness. J Audio Eng Society 45:224–240

Moore BCJ, Glasberg BR (2004) A revised model of loudness perception applied to cochlear hearing loss. Hear Res 188:70–88

Moore BCJ, Glasberg BR (2007) Modelling binaural loudness. J Acoust Soc Am 121:1604–1612

Reynolds GS, Stevens SS (1960) Binaural summation of loudness. J Acoust Soc Am 32:1337–1344

Rothauser EH, Chapman WD, Guttman N, Nordby KS, Silbiger HR, Urbanek GE, Weinstock M (1969) I.E.E.E. recommended practice for speech quality measurements. IEEE Trans Audio Electroacoust 17:227–246

Sivonen VP, Ellermeier W (2006) Directional loudness in an anechoic sound field, head-related transfer functions, and binaural summation. J Acoust Soc Am 119:2965–2980

Whilby S, Florentine M, Wagner E, 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, Scharf B (1965) A model of loudness summation. Psychol Rev 72:3–26

Zwicker E, Zwicker UT (1991) Dependence of binaural loudness summation on interaural level differences, spectral distribution and temporal distribution. J Acoust Soc Am 89:756–7645