Loudness constancy for noise and speech: How instructions and source information affect loudness of distant sounds

Gauthier Berthomieu, Vincent Koehl, Mathieu Paquier

Loudness constancy for noise and speech: how instructions and source information affect loudness of distant sounds

Gauthier Berthomieu[1*], Vincent Koehl[1] and Mathieu Paquier[1]

[1]Univ Brest, Lab-STICC, CNRS, UMR 6285, F-29200 Brest, France.

*Corresponding author(s). E-mail(s): gauthier.berthomieu@univ-brest.fr;

Abstract

The physical properties of a sound evolve when traveling away from its source. As an example, the sound pressure level at the listener's ears will vary according to their respective distance and azimuth. However, several studies have reported loudness to remain constant when varying the distance between the source and the listener. This loudness constancy has been reported to occur when the listener focused attention on the sound as emitted by the source (namely the distal stimulus). Instead, the listener can focus on the sound as reaching the ears (namely the proximal stimulus). The instructions given to the listener when assessing loudness can drive focus toward the proximal or distal stimulus. However, focusing on the distal stimulus requires to have sufficient information about the sound source, which could be provided by either the environment or by the stimulus itself. The present study gathers three experiments designed to assess loudness when driving listeners' focus toward the proximal or distal stimuli. Listeners were provided with different quality and quantity of information about the source depending on the environment (visible or hidden sources, free field or reverberant rooms) and on the stimulus itself (noise or speech). The results show that listeners reported constant loudness when

asked to focus on the distal stimulus only, provided enough information about the source was available. These results highlight that loudness relies on the way the listener focuses on the stimuli and emphasize the importance of the instructions that are given in loudness studies.

1 Introduction

1.1 Loudness constancy

Perceptual constancy refers to the tendency to perceive an object as having constant features despite changes in the properties of the presented stimulus [1]. It is known to occur when the observer is familiar with the perceived object [2] and has mainly been investigated in visual perception, in studies about size [3] or shape [4] constancy.

Loudness constancy refers to a situation where the loudness produced by a given sound remains constant despite changes in its physical properties. As an example, the sound level decreases by 6 dB per doubling of the distance in free field. However, loudness constancy has been observed with varying source distance [5–7]. In the same way, loudness constancy has been observed between monaural and binaural presentations of same signals [8] despite the binaural summation process [9]. In these studies, the sounds displayed at the listeners' ears could differ according to the experimental conditions whereas the sounds emitted by the sources were assumed to be constant. In studies focusing on source distance, the sources emitted constant sounds but their varying distance led to stronger at-ear sound pressure levels at the closest source distances. In studies focusing on monaural versus binaural presentations, the signals emitted by the source were constant but the signals reaching the listeners' ears were not, as one ear could be either open or occluded. A distinction can then be made between the sound emitted by the source – the distal stimulus – and

the sound reaching the receiver – the proximal stimulus [10]. In the aforementioned experimental setups, whereas the distal stimuli were constant (the sounds emitted by the source were unaltered by the experimental manipulations), the proximal stimuli depended on the experimental conditions (i.e. the sound source distance reduced the at-ear sound pressure level). The loudness constancy observed in these situations revealed that loudness judgements followed the consistency of the distal stimuli rather than the inconsistency of the proximal stimuli.

The perception of the size of a familiar object can rely on the subject's past experience with this object (or similar objects). Bolles & Bailey [11] gathered size estimates for familiar objects (e.g. an ashtray or a book) that were verbally described first (with no reference to their size), and then visually presented to the subjects. The results highlight a strong correlation between the estimates made in the two presentation methods, suggesting that the size estimate of a familiar object does not exclusively rely on the visual cues to size, but also comes from the subject's past experience and learning. Mohrmann [12] highlighted that loudness constancy was more likely to occur for familiar stimuli (i.e. music or speech) than for non-familiar stimuli (i.e. pure tones or noises). The past experience of the listeners with these familiar stimuli might have allowed them to estimate loudness on the basis of additional cues. As an example, the perceived distance of speech differ according to whether it is whispered or shouted [13]. Such observation might be due to the ability for the listeners to match the speech timbre to a specific vocal effort (and thus a specific source power), which might provide information about the source distance. Such observation could not apply to non-familiar stimuli such as noises or tones for which the timbre do not inform the listeners on their source's power.

The assignation of a sound to a particular sound source forms a perceptual construct named auditory object according to the definition of Bizley and Cohen [14]. A sound can still be heard without being recognized [15] and will subsequently not be assigned to a particular source. Since the "source familiarity" is defined by Philbeck & Mershon [13] as the stored knowledge upon which one might base an estimate of output power, loudness constancy might be achieved only in cases where the listener perceives an auditory object [16].

Since loudness is a subjective experience, it depends on the way a listener interacts with an auditory object. Specifically, the perception of this auditory object can be affected by the way the listener focuses on the stimulus. As an example, in a cocktail party situation, a listener is able to devote more processing resources when focusing on familiar sounds [17].

Several loudness studies use experimental setups that keep a consistent distal stimulus while modifying the proximal stimulus, e.g. by modifying the source location in directional loudness studies [10, 18, 19] or by occluding one of the listener's ears when investigating binaural loudness summation [8, 20]. Loudness estimates could therefore differ whether the listener focuses on the variable proximal stimulus or on the constant distal stimulus, which could account for significant interindividual differences observed in the literature [21].

The listener's focus can be directly driven by explicit instructions toward a particular stimulus [12]. However, the extent to which the listener focuses on the proximal or distal stimulus appears to depend on the stimulus itself [16]. In cases where the stimulus does not provide the listener with intrinsic information about the sound source (i.e. for an anechoic noise or pure tone displayed through headphones with no visual stimulus), loudness and distance estimates rely solely on the at-ear sound pressure level [22], which could be

due to the lack of information provided about the source [16]. In this way, the environment in which the sounds are presented, the quantity and quality of information available about the sound source or its familiarity to the listener are likely to affect the focus.

In studies that investigate relationships between the source position and the loudness of the emitted sound, any information about the source position or power is likely to attract the listener's focus toward the distal stimulus. The reverberant energy depends on the source power and is almost constant with the distance to the source, whereas the direct energy decreases linearly with the square of distance. Thus, reverberation cues can give information about the sound source distance and power [23], allowing loudness constancy across distance [6]. Timbral cues of familiar sounds such as speech and music can inform the listener on the source power [24] and distance [13, 25]. Visual cues provide accurate distance and power information that could affect loudness estimates [26].

1.2 Present experiments

Directional loudness sensitivity (the extent to which a sound reaching a listener from the side is perceived louder than a frontal sound of same pressure level) of narrow-band noises was studied in a previous experiment [27]. This past study was conducted in two separate sessions, one where the sources were visible loudspeakers and one where sounds were displayed through headphones with no visible source. The results differed between the two sessions, reporting lower directional loudness sensitivity when the sounds were displayed by visible loudspeakers. These results led to hypothesizing that the visual cues to the sound sources drove the listener's focus on the distal stimuli, favoring constant (or at least less varying) loudness with the source position. In this previous

study, the listeners were not asked to explicitly focus on the distal stimuli. They made pairwise comparisons between frontal and lateral sounds and were asked to specify which of the two sounds was the louder. No further specification was given about what their judgments should be based on.

Loudness studies that explicitly drive the listener's focus thanks to direct instructions are rather sparse, but highlight a strong influence of instructions on loudness. Zahorik & Wightman [6] collected loudness estimates using "a free-modulus magnitude estimation procedure in which listeners were carefully instructed to make their judgments based on the sound source power", highlighting a strong loudness constancy when instructing the listeners to focus on the distal stimuli. Mohrmann [12] asked listeners to adjust the output levels of two loudspeakers so that "the two sources – or else the two impacts – appeared to be equally loud", as translated by Brunswik [28]. The adjustments made with regard to the "impacts" (the proximal stimuli) were highly dependent on the source distance (and thus on the level of the proximal stimuli), which was not the case for the adjustments made with regard to the sources (the distal stimuli).

Thus, the instructions given to the listeners seem to be able to drive their focus towards the proximal or distal stimuli, and this focus is likely to affect loudness. However, most loudness studies ask the participants to estimate loudness without giving further specifications (e.g. by asking "How loud is this sound?") [16]. The present study compared loudness judgments gathered when listeners focused on either the proximal or distal stimulus. The listeners' focus was driven through explicit instructions and the amount of information about the source was controlled by manipulating auditory (namely reverberation and timbre) and visual cues given to the listeners. Three experiments were set up, in which consistent sound sources were located at different distances from the

listeners so that the proximal stimuli had varying characteristics. The results were then discussed in terms of loudness constancy, which describes whether loudness was based on the constant distal stimuli or on the varying proximal stimuli.

- The first experiment gathered loudness and distance estimates for noise bursts played at several distances from the listening point. The sound source was either visible or hidden and the experiment took place in both reverberant and anechoic environments. The aim was to observe whether visual and auditory information about the source could affect loudness when asking the listeners to focus on the proximal stimulus. Distance estimates were also collected to determine whether a potential effect could be explained by a modification of the perceived sound source distance thanks to visual and auditory cues (through reverberation). The use of noise bursts deprived the listeners of timbral cues to the source power. The hypothesis behind this experiment was that when instructed to focus on the proximal stimuli, listeners would make loudness estimates that are strongly dependent on the at-ear sound pressure level, and thus on the source distance, whatever the quantity and quality of information about the source.
- The second experiment gathered loudness estimates for noise bursts played in the same conditions as in the first experiment, but with explicitly driving the listeners' focus on the distal stimulus. The main hypothesis behind this experiment was that when provided with information about the source (that is when the environment is echoic and/or when the source is visible), listeners that are instructed to focus on the distal stimuli would be able to report constant loudness estimates with source distance.
- The third experiment gathered loudness estimates obtained when focusing on (i) the proximal stimuli, (ii) the distal stimuli for speech spoken at several

distances from the listening point. The speaker could also be either visible or hidden and the stimuli were presented in reverberant or anechoic environments. This experiment aimed at observing the evolution of the loudness estimates with sound source distance in both focus situations in a case where intrinsic source power cues were delivered by the stimuli though timbral and visual cues. Two hypotheses were formulated for this experiment. First, it was hypothesized that listeners would still report loudness estimates that depend on the at-ear pressure level (and thus on the source distance) when instructed to focus on the proximal stimuli. Then, it was hypothesized that when instructed to focus on the distal stimuli, loudness constancy could be achieved even in free field with no visual cue to the source thanks to the timbral cues conveyed by the speech stimuli.

2 Experiment 1: proximal noise

2.1 Material and methods

Studying the relations between the distance of a visible sound source and loudness requires the use of relatively complex experimental setups. As an example, loudspeakers can be moved along a rail by using a motorized mount [29]. This experimental design required additional loudspeakers to display noise at high level during the main loudspeaker movement (which could be rather intrusive) in order to mask it.

In the present study, virtual environments were created. This allowed the experimenters to manipulate the source distance, the source level, the source visibility and the room acoustics independently while providing the participants with realistic audiovisual stimuli. The virtual environments were rendered visually by a Head Mounted Display (HMD, model HTC Vive) and

auditorily through headphones (Sennheiser HD 650). A loudspeaker representing the source was included in the virtual environments in front of the participants.

Three different environments were studied in this first experiment : a free field, a small concert hall and a large sports hall. The concert hall and sports hall were virtual copies of real rooms.

A large panel was displayed in the environments, which could fully hide the sound source if placed between the latter and the participant. This occlusion was exclusively visual and did not modify the acoustic stimuli.

2.1.1 Visual stimuli

The visual virtual environments were created according to the recommendations made by Renner et al. [30]. This consisted in providing binocular disparity, using high quality of graphics, carefully adjusting the virtual camera settings, displaying rich virtual environments containing a regularly structured ground texture and enhancing the user's sense of presence. 3D models were created in a 3D computer graphics software (Blender) and imported in a game engine (Unity), in which the virtual environments were rendered. Three distinct environments were created to provide congruent visual and auditory environments favoring sound externalization [31]. For the two reverberant rooms (the sports hall and the concert hall), the visual virtual environments were rooms having the same size and shape as the real rooms. For the anechoic environment (free field), no wall was displayed and the environment consisted in an infinite ground with a check pattern texture forming squares with a side length of 1 m (the subjects were not informed of the size of the squares). Additional depth information was provided by adding traffic cones each 5 m in the axis of the loudspeaker. Fig. 1 shows the three virtual environments from

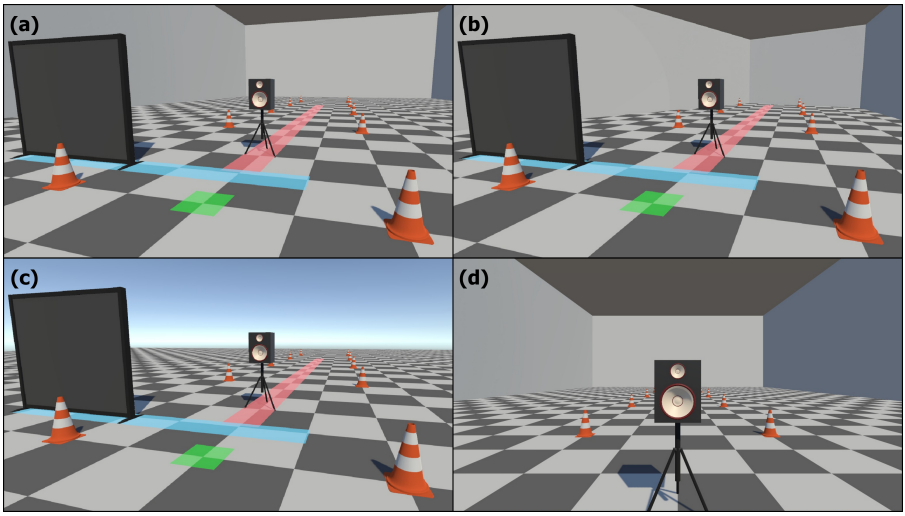

Fig. 1 Virtual environments representing the sports hall (a), the concert hall (b) and the free field (c) as seen from an external point of view. The green square depicts the participant's position within the virtual environment. The panel could move along the blue path to completely hide the source from the participant's position. The source could be placed at several positions along the red path. The square and paths were not displayed during the experiment. (d) shows the source as seen from the participant's position in the sports hall.

an external point of view along with a picture of the sports hall taken from the subjects' point of view. The sound source was depicted as a 3D-rendered loudspeaker as shown in Fig. 1.

2.1.2 Auditory stimuli

During the experiment, binaural stimuli were displayed through headphones. Sound sources were virtually placed at several distances from the participants in the environments. The two reverberant rooms (the sports hall and the concert hall) were captured by measuring Higher Order Ambisonics (HOA) impulse responses of the rooms up to the 4^{th} order. These Spatial Room Impulse Responses (SRIR) were measured by displaying sine sweeps through a loudspeaker (Genelec 8040A) located at 1 m, 2 m, 4 m, 8 m and 16 m from a HOA microphone (Eigenmike EM32). These distances were chosen to provide a large variety of direct-to-reverberant energy ratios in the two rooms and

provide distance doubling so that the direct level decreased linearly with the distance represented on a logarithmic scale. Fig. 2 depicts drawings of the two reverberant rooms with indications on the positions of the loudspeaker and microphone during the recordings. Reverberation times of the two rooms were measured according to ISO 3382-1 [32] specifications and displayed in Table 1.

Environment	T_{30}
Free field	anechoic
Concert hall	0.5 s
Sports hall	2 s

Table 1 Reverberation times (T_{30}) of the three environments.

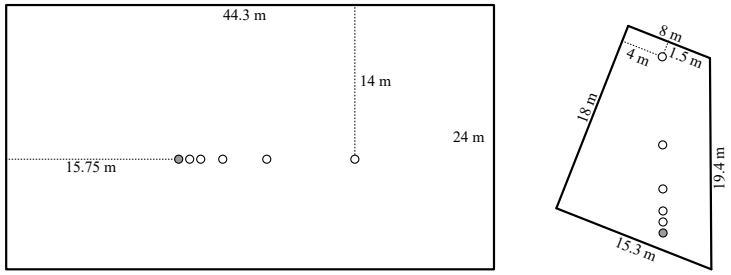

Fig. 2 Drawings of the large sports hall (a.) and small concert hall (b.) in which the recordings were made. The filled circle depicts the microphone position, empty circles depict the loudspeaker positions.

Each SRIR was de-noised using the procedure described by Cabrera et al. [33] and the loudspeaker response was compensated for by using anechoic measurements made by Salmon et al. [34]. A 200 ms frozen white noise was generated in MATLAB and was used as stimulus. It was then convolved with these impulse responses. For the anechoic stimuli, the white noise was directly encoded in ambisonics signals of the 4^{th} order. Head-tracking of the sound source location was carried out thanks to the ambiX Rotator plugin [35] enabling a dynamic restitution of the stimuli. HOA signals were then decoded into binaural signals thanks to the IEM BinauralDecoder plugin, according to

the procedure described by Schörkhuber et al. [36]. The Head-Related Transfer Functions (HRTF) used in the procedure were measured on a dummy head (Neumann KU 100) by Bernschütz [37].

The sounds could be played by the source at three different restitution levels so that one given at-ear pressure level did not match one particular sound source distance. These restitution levels were fixed at 0 dB, −6 dB and −12 dB relatively to a given reference level. This reference level was calibrated by placing the headphones on the dummy head and measuring a sound pressure level of 80 dB SPL at the entrance of the blocked ear canal when the source was at 1 m from the listening point. The dummy head was beforehand calibrated with a sound calibrator (Brüel & Kjær Type 4231) at 1 kHz.

45 stimuli were created, corresponding to the 3 restitution levels for a sound source placed at 5 distances in each of the 3 environments.

2.1.3 Procedure

f (Hz)	31.5	63	125	250	500	1 k	2 k	4 k	8 k	16 k
BN (dB SPL)	43.6	37.0	28.4	19.5	12.7	18.5	13.0	14.2	14.7	14.3
T_{30} (s)		0.28	0.22	0.22	0.15	0.08	0.1	0.14	0.15	

Table 2 Background noise (BN) and reverberation time (T_{30}) measured in the audiometric booth on octave bands of center frequencies f.

The experiment took place in a 2.2 m × 1.85 m × 2.1 m audiometric booth, which background noise BN (RMS, slow [38]) and reverberation time T_{30} measured on octave bands with center frequencies f [39] are displayed in Table 2. where participants sat on a chair. A response interface was available on a tablet computer in front of them. This tablet computer was displayed within the virtual environment by using the HMD built-in front-facing camera. After each response, the sound source was hidden by the large panel before being placed at its next position. A software implemented in Max/MSP ran the procedure

and rendered the audio stimuli in real time, while communicating with Unity thanks to the Open Sound Control (OSC) protocol for gathering head rotation information and for moving the objects within the virtual visual environment.

The experiment was separated in 2 sessions, one gathering egocentric distance estimates and one gathering loudness estimates. Each session consisted of 3 sub-sessions, corresponding to each of the 3 environments. In each sub-session, every distance, restitution level and visibility condition of the sound source (visible or hidden by the panel) was presented 4 times to each participant. This led to 120 trials (5 distances, 3 levels, 2 visibility conditions and 4 repetitions) per participant in each of the 3 sub-sessions (free field, sports hall and concert hall) and in each of the 2 sessions (egocentric distance and loudness estimate). The loudness estimates were gathered using an absolute magnitude estimation protocol [40]. The instructions given to the participants explicitly asked to focus on the proximal stimulus (the sound reaching their ears) by specifying "The louder you hear the sound, the higher the assigned number should be" (translated from french). The egocentric distance was directly estimated in meters. Participants typed their answer on the tablet computer, which was displayed in real time next to the word "Sensation:" (translated from french) for loudness estimates and next to the word "Distance:" (translated from french) for egocentric distance estimates. 20 participants (5 women and 15 men, aged 20 to 25 years) with self-reported normal hearing and normal or corrected to normal vision (the HMD was carefully adjusted so that participants that wear prescription glasses could fit them inside the headset) participated in this experiment and were remunerated for their participation. Each sub-session began with a series of 10 pre-test estimates which responses were not kept. Each session lasted around one hour. Participants attended the two sessions on two separate weeks. One half of the participants began

with the loudness estimate session and the other half began with the distance estimate session. The sub-sessions and the trials within each sub-session were carried out in a random order by each participant. A video recording of the experiment can be found at the following url: https://youtu.be/V1ifR558VO4.

After the experiment, participants filled a 7-item questionnaire inspired by statements from Rébillat et al. [41] in order to evaluate their sensation of presence within the virtual environment. The participants rated each item on a 7-points Likert scale [42]. Each answer was assigned a score between -3 and 3, -3 meaning a poor presence sensation and 3 meaning a strong presence sensation. The scores were averaged across items and participants and showed an overall positive score ($\mu = 0.93$, $\sigma = 0.63$), showing that the sensation of presence was globally satisfying within the virtual environment.

2.2 Results and discussion

2.2.1 Loudness

Loudness estimates were gathered using an absolute magnitude estimation protocol. These results were subsequently normalized across participants prior to statistical analysis, according to the procedure described by Altmann et al. [7] :

1) Each single estimate was converted to its logarithm. **2)** The arithmetic mean x_{sc} of these logarithm estimates was computed for each subject s and for each experimental condition c (across repetitions). **3)** From each x_{sc} value, a normalized n_{sc} value was obtained : $n_{sc} = x_{sc} - X_s + X$, where X_s is the arithmetic mean of the logarithmic estimates of subject s across conditions and X the grand mean of the logarithmic estimates for all subjects and all conditions. X enabled to depict loudness estimates that are in the same order

of magnitude as the participants' estimates in the figures.

A repeated-measures ANalysis Of VAriance (ANOVA) was performed on the normalized logarithms n_{sc}. This analysis included 4 factors: the environment (3 levels), the source distance (5 levels), the source level (3 levels) and the source visibility (2 levels). The residuals of the linear model were normally distributed. The results of the ANOVA are presented in Table 3.

Loudness estimates made while focusing on the proximal stimuli

Cases	SS	DF	MS	F	Sig.p
E	1.783	2	.892	9.879	**< .001**
D	36.354	1.125*	32.316	186.971	**< .001**
L	24.764	1.042*	23.759	217.212	**< .001**
V	.017	1	.017	2.125	.161
E × D	2.059	2.883*	.714	37.202	**< .001**
E × L	.055	2.451*	.022	1.492	.233
D × L	.269	4.077*	.066	5.959	**< .001**
E × V	.004	2	.002	.528	.594
D × V	.014	2.362*	.006	.712	.518
L × V	.009	2	.004	1.230	.304
E × D × L	.051	6.806*	.007	1.004	.430
E × D × V	.010	4.611*	.002	.408	.828
E × L × V	.024	4	.006	1.888	.121
D × L × V	.025	4.379*	.006	.838	.514
E × D × L × V	.058	6.429*	.009	1.219	.299

* The degree of freedom was adjusted with a Greenhouse-Geisser correction following a violation of the assumption of sphericity.

Table 3 Results of the ANOVA conducted on the normalized logarithms of the loudness estimates of noise made while focusing on the proximal stimuli. E stands for the Environment, D for the virtual source Distance, L for the source Level and V for the source Visibility.

On the one hand, the visibility factor did not prove to have any simple effect or interaction effect on loudness. On the other hand, the distance factor proved to have on significant effect on loudness ($\mathrm{F}(1.125^{1}, 21.374) = 186.971, p < .001$) and significantly interacted with the environment factor ($\mathrm{F}(2.883^{1}, 54.774) = 37.202$, $p < .001$). As can be noted in Fig. 3 which depicts[2] the loudness as a function of the distance for each of the 3 environments, loudness decreased

[1]The degree of freedom was adjusted with a Greenhouse-Geisser correction because the sphericity assumption of the results was violated.

[2]Here and for the rest of the paper, the results are displayed with geometric means (i.e. the inverse logarithm of the n_{sc} values).

with the distance.

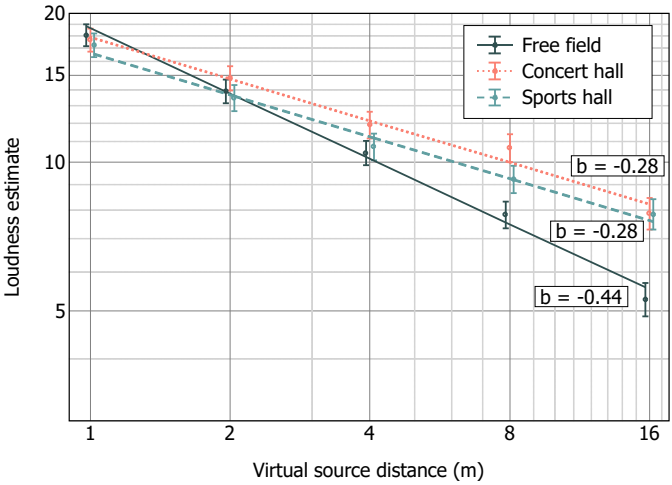

Fig. 3 Geometric mean loudness estimate made when listeners were instructed to focus on the proximal stimuli, as a function of the sound source distance in free field (solid line), in the concert hall (dotted line) and in the sports hall (dashed line), with 95% confidence intervals. The b exponent next to each curve depicts the steepness of the slope.

The loudness-distance functions depicted in Fig. 3 were obtained thanks to a fitting of the data to power functions:

$$L = k \cdot r^b \tag{1}$$

where L is the loudness, r the distance and k a constant. The b exponent indicated in Eq. 1 represents the slope of the loudness-distance function and is indicated next to its respective function in Fig. 3. A perfect loudness constancy would thus be indicated by an exponent $b = 0$. The power fittings obtained in the same way by Zahorik & Wightman [6] assumed loudness constancy of broadband noise up to $|b| \simeq 0.1$. The only fitting that assumed failure of loudness constancy in this study revealed $b = -0.35$.

The exponents obtained in the present experiment were $b = -0.28$ in both reverberant environments and $b = -0.44$ in the anechoic environment. Therefore, the results do not show loudness constancy with sound source distance when listeners were asked to focus on the proximal stimulus in any of the three environments under investigation. The shallower slope in the reverberant environments is likely to be caused by the reverberant energy, which made the overall at-ear level to be higher in these environments than in the anechoic one for distant stimuli. Indeed, while the at-ear level decreased by 6 dB each doubling of the source distance in free field, the isotropy of the reverberant energy caused the at-ear level to decrease by less than 6 dB each doubling of the source distance in the echoic rooms.

2.2.2 Perceived distance

The distance estimates were gathered in order to explain potentials effects of visibility on the loudness estimates. Since distance estimates are approximately normally distributed along a logarithmic scale [43], each distance estimate was converted to its logarithm before being statistically analyzed. The values were not normalized across participants, since each participant was assumed to use the same scale (meter unit). A repeated-measures ANOVA was conducted on the distance estimates, including the same factors as the one performed on the loudness results: the environment (3 levels), the source distance (5 levels), the source level (3 levels) and the source visibility (2 levels). The residuals of the linear model were normally distributed. The results of the ANOVA are presented in Table 4.

The distance estimates varied differently with the virtual source distance in the three environments. Fig. 4 shows that distance estimates were overall smaller in the anechoic environment than in the reverberant environments,

Distance estimates					
Cases	**SS**	**DF**	**MS**	**F**	**Sig.p**
E	9.165	1.132*	8.097	14.602	**.001**
D	163.111	1.136*	143.526	244.124	**< .001**
L	16.564	1.183*	14.002	235.614	**< .001**
V	.170	1	.170	3.543	.075
E × D	.693	2.423*	.286	3.948	**.020**
E × L	.807	2.070*	.390	11.130	**< .001**
D × L	1.799	2.562*	.702	21.463	**< .001**
E × V	.294	1.374*	.214	7.659	**.006**
D × V	1.531	1.610*	.951	13.737	**< .001**
L × V	1.476	1.075*	1.373	14.996	**< .001**
E × D × L	.196	4.846*	.040	1.525	.191
E × D × V	.062	8	.008	.910	.510
E × L × V	.005	2.595*	.002	.170	.893
D × L × V	.098	3.446*	.028	1.631	.184
E × D × L × V	.091	7.430*	.012	.848	.556

* The degree of freedom was adjusted with a Greenhouse-Geisser correction following a violation of the assumption of sphericity.

Table 4 Results of the ANOVA conducted on the logarithms of the distance estimates. E stands for the Environment, D for the virtual source Distance, L for the source Level and V for the source Visibility.

as confirmed by the simple environment effect ($F(1.132^1, 21.505) = 14.602$, $p < .001$), and that distance estimates increased more steeply with distance in the sports hall than in the concert hall, as confirmed by the interaction between distance and environment ($F(2.423^1, 46.037) = 3.948$, $p = .02$).

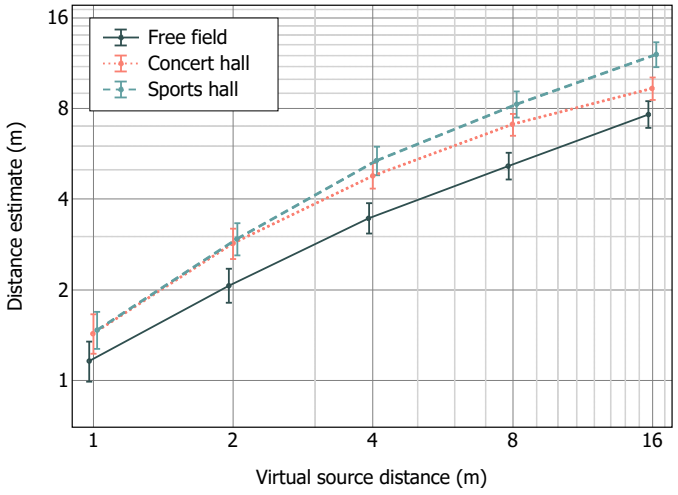

Fig. 4 Geometric mean distance estimate as a function of the virtual sound source distance in free field (solid line), in the concert hall (dotted line) and in the sports hall (dashed line), with 95% confidence intervals.

The visibility factor did not prove to have a significant effect on distance estimates, but significantly interacted with the three other factors. Distance estimates were closer to the actual distance values when the source was visible than when it was hidden (Fig. 5), as confirmed by the interaction between distance and visibility ($F(1.61^1, 30.581) = 13.737$, $p < .001$). Distance estimates depended less on the sound source level when the latter was visible that when it was hidden (not shown here), as confirmed by the interaction between level and visibility ($F(1.075^1, 20.424) = 14.996$, $p < .001$). Distance estimates were less different across environments when the source was visible than when it was hidden (not shown here), as confirmed by the interaction between environment and visibility ($F(1.374^1, 26.097) = 7.659$, $p = .006$). These interactions are in agreement with the literature and show an increase in accuracy when visual cues are available [29]. In the presence of visual cues, which provided the participants with absolute and accurate distance cues [44], the relative distance cues (such as the source level [45]) or the less accurate distance cues (such as the direct-to-reverberant energy ratio) then had a weaker influence on distance estimates than when the source was hidden.

2.2.3 Discussion

While the perceived distance of the sound source depended on its visibility, the estimates made with and without visual cues to the source appeared to be closer from each other than what could be expected based on the literature. Listeners usually tend to overestimate the distance to sources closer than 1 m and to underestimate that of remote sources [43]. While these biases can be observed in Fig. 5, the source distance beyond which listeners underestimated the distance was about 4 m. While such biases were not reported to occur in

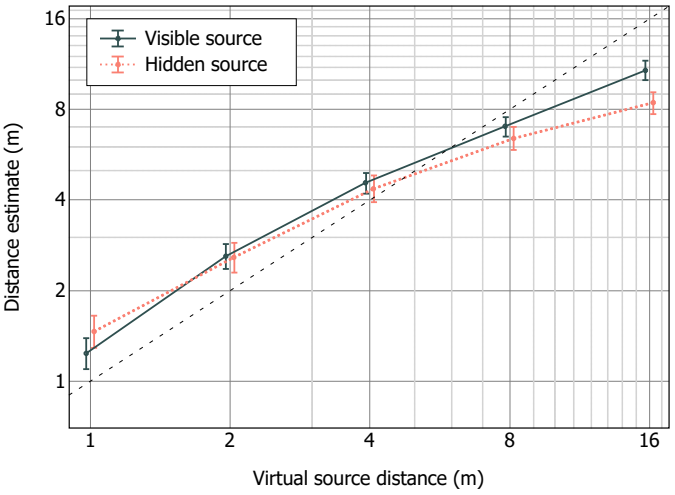

Fig. 5 Geometric mean distance estimate as a function of the virtual sound source distance for a visible (solid line) or hidden (dotted line) source, with 95% confidence intervals, along with the actual source distance (dashed line).

visual depth perception [46], Fig. 5 shows a relatively small but existing overestimation of close visible source distances and an underestimation of farther visible source distances. Lastly, auditory distance estimates usually tend to be more variable than visual distance estimates [45], while Fig. 5 exhibits similar variances in distance estimates to visible or hidden sources. However, the comparisons made in the literature usually involve distance estimates made in separate studies, with either visible or hidden sources (but not both). In the present study, estimates were gathered within blocks mixing both visible and hidden sources. The observed better-than-usual performances in auditory distance estimate might then have resulted from a carryover across visible and hidden source presentations, where participants could have matched the (precise) visually-determined distances of audiovisual stimuli with the auditory stimuli, enhancing their performances when presented with the auditory stimuli only. Similar carryover have been reported in the literature [47] between distance estimates made for visual or auditory targets in separate blocks.

However, the results of this experiment show that the loudness estimates of white noise bursts gathered when asking the listeners to focus on the proximal stimuli were identical whether the sound source was visible or hidden in every experimental conditions under investigation, while the distance estimates statistically depended on the source visibility. The loudness of the white noises displayed by the source thus followed the at-ear sound pressure level decrease, regardless of the distance at which the sound source was perceived. These results are in accordance with previous studies that did not find any loudness constancy with source distance for non-familiar stimuli [7, 48], for which the perceptual construct of auditory object (and therefore of source distance) may be irrelevant. Consequently, the following experiment will study interactions between visual distance cues and loudness estimate when asking the listeners to focus on the distal stimulus, in both reverberant and anechoic environments.

3 Experiment 2: distal noise

3.1 Materials and methods

The protocol of this experiment was similar to that of the experiment 1 (see section 2.1). The same stimulus was used and was displayed either in an anechoic environment (free field), or in the most reverberant environment from experiment 1 (the large sports hall). Thus, 30 auditory stimuli were created, corresponding to a white noise displayed at 3 restitution levels by a sound source placed at 5 different distances in 2 environments (anechoic or reverberant).

3.1.1 Procedure

This experiment aimed at collecting loudness estimates by using a similar procedure as the one used in the loudness session of experiment 1. The experiment was done in one session consisting of 2 sub-sessions corresponding to each of the 2 environments. 120 trials (5 distances, 3 levels, 2 visibility conditions and 4 repetitions) were carried out in random order by each participant within each of the 2 sub-sessions (free field and sports hall). The loudness estimates were obtained by using the same absolute magnitude estimation protocol as in experiment 1. The participant's attention was explicitly focused on the distal stimulus by specifying "The louder the sound is played by the source, the higher the assigned number should be" (translated from french). Participants typed their answer on the tablet computer. Since the strength of the sound displayed by a source relates to its power, their answer was displayed in real time next to the word "Power:" (translated from french). 17 participants (3 women and 14 men, aged 20 to 27 years) with self-reported normal hearing and normal or corrected to normal vision participated in this experiment and were remunerated for their participation. They were not involved in experiment 1. Each session lasted about 45 minutes.

3.2 Results and discussion

The loudness estimates were normalized the same way as the loudness estimates gathered in the experiment 1 (see section 2.2).

A repeated-measures ANOVA was conducted on the loudness estimates obtained in this experiment. 4 factors were included: the environment (2 levels), the distance (5 levels), the sound source level (3 levels) and the source visibility (2 levels). The residuals of the linear model were normally distributed. The results of the ANOVA are presented in Table 5.

Loudness estimates made while focusing on the distal stimuli

Cases	SS	DF	MS	F	Sig.p
E	.007	1	.007	.037	.851
D	3.240	1.091*	2.971	5.938	**.024**
L	11.049	1.094*	10.103	71.034	**< .001**
V	.003	1	.003	.167	.688
E × D	2.340	4	.585	60.862	**< .001**
E × L	.119	2	.059	4.533	**.018**
D × L	.036	3.204*	.011	.589	.636
E × V	.217	1	.217	10.666	**.005**
D × V	1.634	1.647*	.992	21.641	**< .001**
L × V	.001	2	¡.001	.118	.889
E × D × L	.040	3.342*	.012	.892	.460
E × D × V	.369	2.422*	.152	14.645	**< .001**
E × L × V	.004	2	.002	.325	.725
D × L × V	.126	3.773*	.034	3.878	**.008**
E × D × L × V	.025	8	.003	.865	.548

* The degree of freedom was adjusted with a Greenhouse-Geisser correction following a violation of the assumption of sphericity.

Table 5 Results of the ANOVA conducted on the normalized logarithms of the loudness estimates of noise made while focusing on the distal stimuli. E stands for the Environment, D for the virtual source Distance, L for the source Level and V for the source Visibility.

Loudness-distance functions obtained in further analyses were subsequently fitted to power functions as described in Eq. 1. The corresponding b exponents are indicated next to their respective functions in the following figures.

3.2.1 Distance

The distance factor proved to have a significant effect on the loudness ($\mathrm{F}(1.091^1, 17.449) = 5.938$, $p = .024$), which does not support the loudness constancy hypothesis for the overall results. However, several significant interactions involving the distance factor give more in-depth information about the loudness constancy of these estimates.

3.2.2 Environment × Distance

The loudness estimates did not depend on the distance in the same way whether the environment was anechoic or reverberant ($\mathrm{F}(4, 64) = 60.862$, $p < .001$). As can be seen in Fig. 6, loudness estimates were less dependent on the sound source distance in the sports hall (the reverberant environment,

in dotted line, with an exponent $b = -0.02$) than in free field (anechoic environment, in solid line, with an exponent $b = -0.24$). This observation is in accordance with the hypothesis made by Zahorik & Wightman [6], as loudness estimates gathered while focusing on the distal stimulus are less dependent on sound source distance (and thus on the at-ear sound level) when power and distance cues supplied by the reverberant field are available to the listeners.

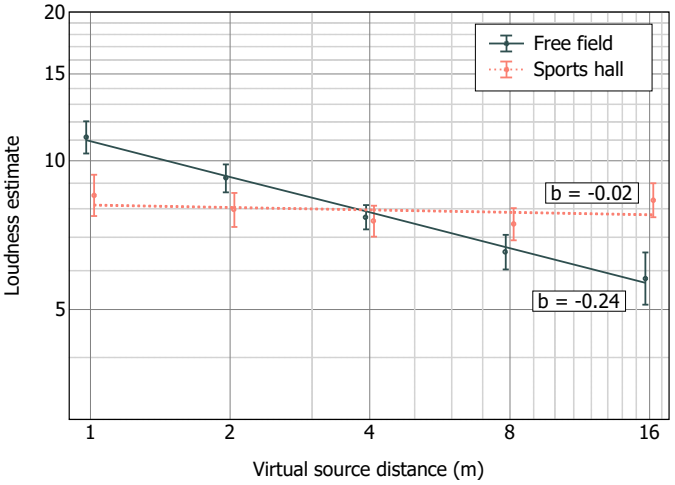

Fig. 6 Geometric mean loudness estimate made when listeners were instructed to focus on the distal stimuli, as a function of the sound source distance in free field (solid line) and the sports hall (dotted line), with 95% confidence intervals.

3.2.3 Distance × Visibility

The loudness estimates did not depend on the distance in the same way whether the sound source was visible or hidden ($F(1.647^1, 26.356) = 21.641$, $p < .001$). As can be seen in Fig. 7, the loudness was less dependent on the source distance when it was visible (solid line, with an exponent $b = -0.04$) than when it was hidden (dotted line, with an exponent $b = -0.22$). The comparison between Fig. 6 and Fig. 7 and between the b exponents obtained on

each loudness-distance function shows that the reverberant field and the visibility affected loudness estimate in a similar way when listeners focused on the distal stimulus.

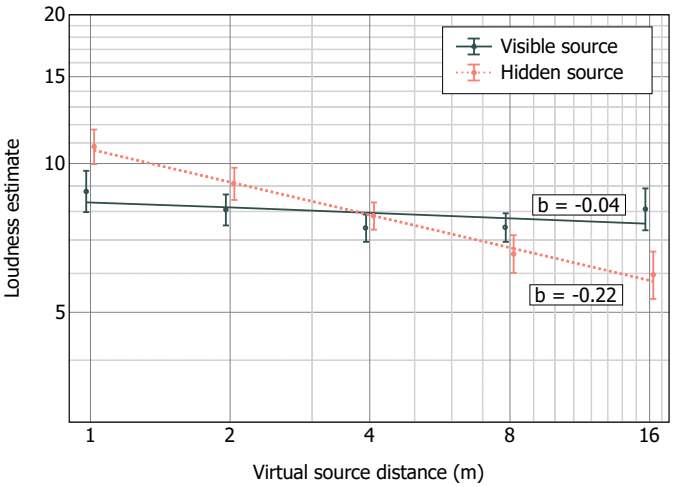

Fig. 7 Geometric mean loudness estimate made when listeners were instructed to focus on the distal stimuli, as a function of the sound source distance when the source was visible (solid line) or hidden (dotted line), with 95% confidence intervals.

3.2.4 Environment × Distance × Visibility

This significant interaction ($F(2.422^1, 38.759) = 14.645$, $p < .001$) is illustrated in Fig. 8. Loudness estimates made in the anechoic environment (a) and in the sports hall (b) are depicted as a function of the sound source distance when the latter was visible (solid line) or hidden (dotted line).

The power fitting made on the loudness estimates gathered in the anechoic environment when the source was hidden reveals an exponent $b = -0.37$. Thus, the obtained loudness-distance functions had a similar slope as the loudness-distance functions obtained in experiment 1 where listeners focused on the proximal stimulus (Fig. 3). This is due to the only cue available to

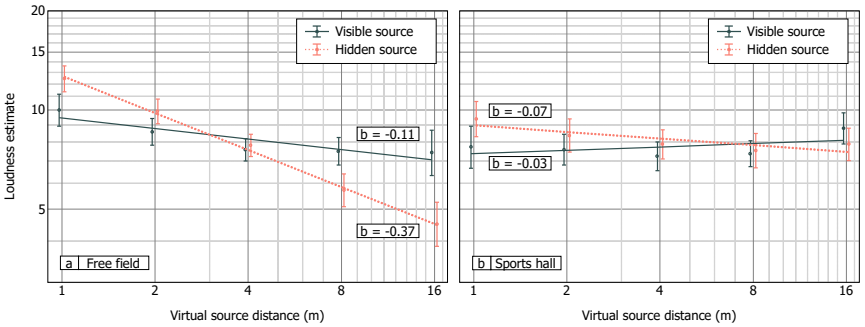

Fig. 8 Geometric mean loudness estimate made when listeners were instructed to focus on the distal stimuli, as a function of the sound source distance in free field (a) and the sports hall (b) when the source was visible (solid line) or hidden (dotted line), with 95% confidence intervals. The b exponent next to each curve depicts the steepness of the slope.

the participants being the at-ear sound level, which depended on the source distance.

In the reverberant environment, the exponents $b = -0.07$ (hidden source) and $b = 0.03$ (visible source) are both close to zero. In the anechoic environment, the estimates depended less on the sound source distance when the source was visible than when it was hidden. The visible source led to an exponent $b = -0.11$, which is slightly higher (in absolute value) than the b exponents obtained by Zahorik & Wightman [6]. However, Bonferroni post-hoc tests showed that each pairwise comparison of the results obtained at the five distances when the source as visible ($b = -0.11$) revealed no significant difference. In comparison, each pairwise comparison for the results obtained when the source was hidden in this environment ($b = -0.37$) led to a significant difference. Cohen's d were computed on these significant differences, highlighting medium to large effects ($d = 0.681$ for the smallest effect, which was found between the results obtained at 1 m and 2 m).

3.2.5 Discussion

The results of this experiment show that when focusing on the distal stimulus:

- The loudness estimates gathered in the anechoic environment were constant with the source distance when the latter was visible.
- The loudness estimates gathered in the reverberant environment were constant with the source distance whether the latter was visible or hidden.
- The reverberant field (which provides the participant with distance and power information) and the visual cues (which provide the participant with distance information) both led to a similar extent of loudness constancy of white noise bursts with their source distance.

In agreement with the results obtained by Zahorik & Wightman [6], the present experiment observes loudness estimates that do not depend on the source distance in a reverberant environment when listeners were asked to focus on the distal stimulus. According to the authors, the loudness estimates might be based on the reverberant energy, which provides the listener with direct information about the sound source power. The results of the present experiment do not contradict this hypothesis but also reveal a loudness constancy with source distance in an anechoic environment, provided that the sound source was visible. This observation highlights the ability to gather the perceived at-ear sound level and distance in order to estimate the loudness of a distal stimulus. According to this statement, listeners might be able to accurately deduct to what extent the source is powerful even when they hear a quiet sound but see a distant source.

The following experiment will study loudness of speech in anechoic and reverberant environments when listeners are asked to focus on the proximal or distal stimulus. This sound signal, contrarily to the white noise studied so far, provides the listener with information about the sound source power (through the vocal effort), which might affect the relationships between loudness and distance [49, 50].

4 Experiment 3: proximal and distal speech

4.1 Materials and methods

The protocol of this experiment was close to that of experiments 1 and 2 (see section 2.1). However, the auditory and visual stimuli were not the same. Here, a speaker pronounced words with several vocal efforts whereas in the precedent experiments a loudspeaker displayed white noise bursts at several levels.

4.1.1 Auditory stimuli

The environments under investigation were the same as in experiment 2 (the free field and the sports hall). The auditory stimuli used in this experiment were words pronounced by a speaker. These words were recorded in a recording booth by an omnidirectional microphone (DPA 4006A) placed close to the speaker's mouth (15 cm). During the recording, the speaker had to pronounce the words at 3 different levels, which produced 68 dB SPL, 74 dB SPL and 80 dB SPL at 1 m from him, creating stimuli with 3 different vocal efforts (producing the same levels as the 3 restitution levels of experiments 1 and 2). A reference omnidirectional microphone (DPA 4006A) was placed at 1 m from the listener's mouth in the axis of the recording microphone. This microphone was calibrated with a sound calibrator (Brüel & Kjær Type 4231) at 1 kHz. The output of this microphone was analyzed in a software implemented in Max/MSP. The sound pressure level was computed on a 1 s-long sliding window (with 50 ms-long steps) and a graphical interface provided the speaker with a real-time indication on whether the words had been pronounced at the desired level or not. The level was considered as correct if it was measured within a ± 1 dB tolerance margin. Moreover, the dynamic range of the pronounced words was measured and needed to be ≤ 10 dB for the recorded words to be considered valid. 3 disyllabic french words close to the spondees

used by Epstein and Florentine [8] were recorded so that different spectral contents were taken into account. These 3 words, "Réchaud", "Normand" and "Caveau", were picked in disyllabic lists used in vocal audiometry [51]. The speaker was then virtually positioned in the virtual environments thanks to the same HOA encoding process as explained in the description of experiment 1, at 1, 2, 4, 8 and 16 m from the listening point. 90 auditory stimuli were created this way, corresponding to each of the 3 words spoken at 3 different levels from 5 different distances in the 2 environments.

4.1.2 Visual stimuli

Expression of the face and body of a speaker depends on the produced vocal effort, which could have an effect on the perceived loudness of speech [24]. A visual representation of the speaker was therefore displayed in the virtual environment. The speaker was filmed (Blackmagic URSA Mini Pro camera) in front of a blue screen while he pronounced the 3 words at the 3 required vocal efforts. 9 videos were recorded this way, corresponding to each of the 3 words pronounced with the 3 vocal efforts. The videos of the speaker were then displayed in the virtual environment at the required distance from the listening point (see Fig. 9), synchronized with the auditory stimuli.

4.1.3 Procedure

This experiment collected loudness estimates by using a similar procedure as the one used in the loudness session of experiment 1 and in experiment 2 (see section 2.1). The experiment was separated in 2 sessions, one where listeners were asked to focus on the proximal stimuli and one where listeners were asked to focus on the distal stimuli. Each session consisted in 2 subsessions, corresponding to each of the 2 environments (the sports hall and

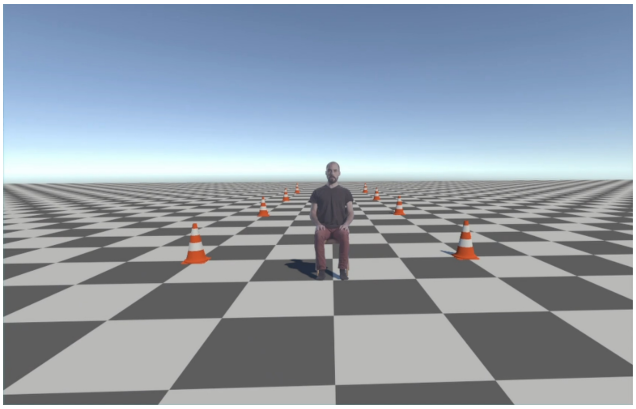

Fig. 9 Speaker located 4 m away from the participant's position within a visual environment.

the free field). In each sub-session, every distance, word, restitution level and visibility condition of the speaker was presented 2 times to each participant in a random order. This led to 180 trials (5 distances, 3 words, 3 levels, 2 visibility conditions and 2 repetitions) per participant in each of the 2 sub-sessions (free field and sports hall) of the 2 sessions (focus on the proximal or on the distal stimuli). Both estimates were obtained by using an absolute magnitude estimation protocol. In the former session, the instructions given to the participants explicitly asked to focus on the proximal stimuli by specifying "The louder you hear the sound, the higher the assigned number should be" (translated from french). In the latter session, the participants' attention was explicitly focused on the distal stimuli by specifying "The louder the person spoke, the higher the assigned number should be" (translated from french). Participants typed their answer on the tablet computer, which was displayed in real time next to the word "Sensation:" (translated from french) for loudness estimates gathered when focusing on the proximal stimuli and next to the word "Vocal effort:" (translated from french) for loudness estimates gathered when focusing on the distal stimuli. 17 participants (6 women and 11 men, aged 20 to 26 years) with self-reported normal hearing and normal or corrected to

normal vision participated in the experiment and were remunerated for their participation. They were not involved in experiments 1 and 2. Each session lasted around one hour. A video recording of the experiment can be found at the following url: https://youtu.be/F3xAx_j0YZw.

4.2 Results and discussion about loudness estimates gathered when focusing on the proximal stimuli

Loudness estimates were normalized according to the same procedure as in experiments 1 and 2 (see section 2.2). A repeated-measures ANOVA was conducted on the normalized logarithms of the results. 5 factors were included in this analysis: the environment (2 levels), the speaker distance (5 levels), the pronounced word (3 levels), the word production level (3 levels) and the speaker visibility (2 levels). The residuals of the linear model were normally distributed. The results of the ANOVA are presented in Table 6.

4.2.1 Visibility

The visibility factor proved to have a significant effect on the loudness estimates ($F(1, 16) = 12.613$, $p = .003$). As can be seen in Fig. 10, the overall estimates were slightly larger when the speaker was hidden than when he was visible. The visibility factor did not significantly interact with any of the other factors.

4.2.2 Distance

The loudness estimates were globally larger when the speaker was close to the participants than when he was remote (not shown here) as confirmed by the significant effect of the distance factor ($F(1.096^1, 17.541) = 69.480$, $p < .001$).

Loudness estimates made while focusing on the proximal stimuli

Cases	SS	DF	MS	F	Sig.p
E	9.873	1	9.873	$1.867 \cdot 10^{1}0$	**< .001**
D	44.402	1.096*	40.502	69.480	**< .001**
W	.214	2	.107	17.132	**< .001**
L	21.653	1.090*	19.872	86.924	**< .001**
V	0.59	1	.059	12.613	**.003**
E × D	4.750	1.493*	3.181	39.470	**< .001**
E × W	.063	2	.031	5.831	**.007**
E × L	.255	2	.128	9.275	**< .001**
E × V	.004	1	.004	.338	.569
D × W	.101	4.444*	.023	2.013	.095
D × L	.310	3.603*	.086	3.071	**.027**
D × V	.032	4	.008	1.016	.406
W × L	.036	4	.009	1.173	.331
W × V	.025	2	.012	3.224	.053
L × V	.001	2	.001	.136	.873
E × D × W	.064	3.522*	.018	1.105	.360
E × D × L	.084	3.953*	.021	1.422	.237
E × D × V	.059	4	.015	1.467	.223
E × W × L	.047	2.542*	.018	2.186	.114
E × W × V	.019	2	.009	2.155	.132
E × L × V	.006	1.366*	.005	.528	.530
D × W × L	.124	6.611*	.019	1.205	.308
D × W × V	.059	3.934*	.015	1.228	.308
D × L × V	.055	3.795*	.014	.727	.570
W × L × V	.012	4	.003	.520	.721
E × D × W × L	.058	5.775*	.010	.677	.663
E × D × W × V	.031	3.884*	.008	.558	.689
E × D × L × V	.037	3.966*	.009	.787	.537
E × W × L × V	.005	4	.001	.266	.899
D × W × L × V	.056	4.916*	.011	.453	.807
E × D × W × L × V	.133	6.106*	.022	1.265	.280

* The degree of freedom was adjusted with a Greenhouse-Geisser correction following a violation of the assumption of sphericity.

Table 6 Results of the ANOVA conducted on the normalized logarithms of the loudness estimates of speech made while focusing on the proximal stimuli. E stands for the Environment, D for the virtual source Distance, W for the spoken Word, L for the source Level and V for the source Visibility.

4.2.3 Environment × Distance

The loudness estimates did not depend on the speaker distance in the same way in the two environments under investigation ($F(1.493^1, 23.896) = 39.470$, $p < .001$). The loudness-distance functions related to these two environments were fitted to power functions as described in Eq. 1. The corresponding b exponents are indicated next to their respective functions depicted in Fig. 11.

The two b exponents revealed by the power fittings ($b = -0.37$ in the anechoic environment and $b = -0.2$ in the reverberant environment) are higher

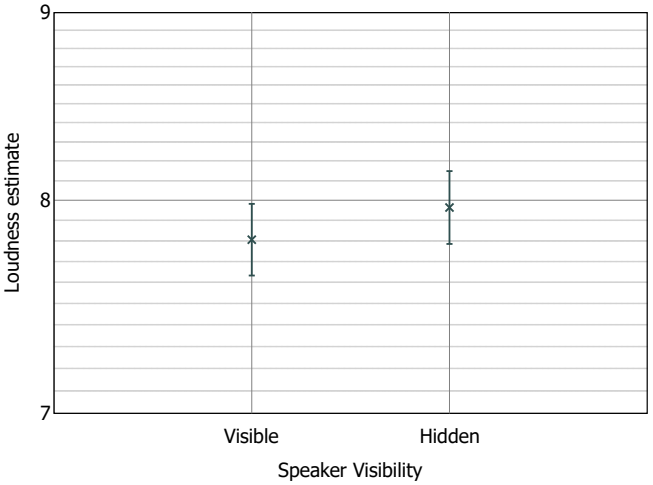

Fig. 10 Geometric mean loudness estimate made when listeners were instructed to focus on the proximal stimuli, as a function of the speaker visibility, with 95% confidence intervals.

(in absolute values) than 0.1 and the estimates should thus not be considered constant with distance according to the literature [6], even if the loudness did not depend on the speaker distance in the same way in the two environments. The differences observed in the two environments are likely to be explained by the at-ear sound level. The latter was stronger in the sports hall than in free field at remote distances, since the reverberant energy decreases softly with the distance.

4.2.4 Discussion

The loudness of speech was stronger when the speaker was hidden than when he was visible. However, this factor did not significantly interact with the distance of the speaker. This suggests that the loudness did not depend on the distance at which the speaker was perceived, which is supported by the results of Experiment 1 (see 2.2).

In the two visibility conditions, the loudness decreased with the sound source distance. This decrease was about the same as the one observed for

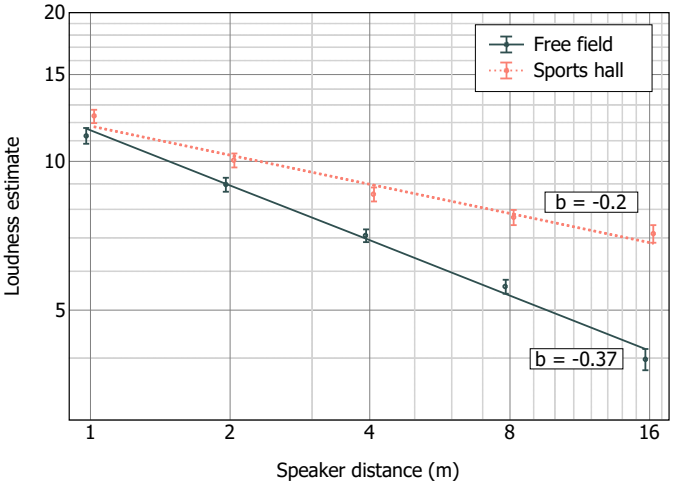

Fig. 11 Geometric mean loudness estimate made when listeners were instructed to focus on the proximal stimuli, as a function of the speaker distance in free field (solid line) and in the sports hall (dashed line), with 95% confidence intervals. The *b* exponent next to each curve depicts the steepness of the slope.

white noises when asking the listeners to focus on the proximal stimuli in the experiment 1. The loudness estimates were thereby directed by the at-ear sound level, which varied with the sound source distance, and were not affected by the timbre and perceived vocal effort. There is a discrepancy between this result and those of Pollack [52] and Warren [50], according to which the loudness of speech depended less on the at-ear level than the loudness of non-familiar stimuli such as the white noise used in the experiment 1. The experimental protocols and the instructions given to the participants were yet different from those of the present experiment by several points:

- The participants' focus was not explicitly led towards the proximal stimulus. The instructions given by Warren [50] were notably "What number would you use to describe the loudness of the fainter sound?", without further specifications.

- The participants in the study of Pollack [52] were asked to adjust the output level of a signal so that its loudness reached the loudness (or half the loudness, or twice the loudness) of a reference signal. The participants in the study of Warren [50] had to estimate the loudness of a signal by assigning a score relative to that of an identical signal displayed at a stronger level, which score was fixed to 100. These two methods gathered relative loudness estimates, whereas the present experiment gathered absolute loudness estimates.

4.3 Results and discussion about loudness estimates gathered when focusing on the distal stimuli

The loudness estimates gathered when focusing on the distal stimuli were normalized and analyzed in the same way as the estimates gathered when focusing on the proximal stimuli (see section 4.2). 5 factors were included in the ANOVA: the environment (2 levels), the speaker distance (5 levels), the pronounced word (3 levels), the word production level (3 levels) and the speaker visibility (2 levels). The residuals of the linear model were normally distributed. The results of the ANOVA are presented in Table 7.

4.3.1 Distance

The overall loudness was significantly stronger when the speaker was close to the participants than when he was remote ($F(1.689^1, 27.024) = 29.126$, $p < .001$). The evolution of the estimates with the source distance will however be analyzed in depth by looking at its interaction with visibility.

Loudness estimates made while focusing on the distal stimuli					
Cases	**SS**	**DF**	**MS**	**F**	**Sig.p**
E	.279	1	.279	$3.542 \cdot 10^8$	**< .001**
D	2.344	1.689*	1.388	29.126	**< .001**
W	1.131	1.414*	.800	17.106	**< .001**
L	45.140	1.053*	42.861	104.642	**< .001**
V	.011	1.000	.011	1.051	.321
E × D	.043	1.811	.024	.909	.405
E × W	.007	2	.003	.503	.610
E × L	.167	1.309*	.128	3.669	.059
E × V	¡.001	1	¡.001	.030	.864
D × W	.056	4.595*	.012	2.034	.089
D × L	.041	2.873*	.014	.695	.554
D × V	.103	4	.026	6.535	**< .001**
W × L	3.003	1.591*	1.887	31.807	**< .001**
W × V	.047	2	.024	4.630	**.017**
L × V	.038	1.433*	.027	3.099	.079
E × D × W	.051	8	.006	1.245	.278
E × D × L	.027	3.578*	.008	.761	.542
E × D × V	.026	4	.007	2.110	.090
E × W × L	.020	2.153*	.009	.489	.631
E × W × V	.008	1.442*	.005	.576	.516
E × L × V	.006	2	.003	.663	.522
D × W × L	.062	5.623*	.011	.851	.528
D × W × V	.055	3.258*	.017	1.329	.274
D × L × V	.100	4.704*	.021	3.015	**.017**
W × L × V	.012	2.346*	.005	.431	.685
E × D × W × L	.062	5.886*	.010	.789	.579
E × D × W × V	.040	3.438	.012	.991	.412
E × D × L × V	.020	3.314*	.006	.411	.765
E × W × L × V	.060	2.726*	.022	4.206	**.013**
D × W × L × V	.071	4.898*	.014	.730	.601
E × D × W × L × V	.102	5.839*	.018	1.367	.237

* The degree of freedom was adjusted with a Greenhouse-Geisser correction following a violation of the assumption of sphericity.

Table 7 Results of the ANOVA conducted on the normalized logarithms of the loudness estimates of speech made while focusing on the distal stimuli. E stands for the Environment, D for the virtual source Distance, W for the spoken Word, L for the source Level and V for the source Visibility.

4.3.2 Distance × Visibility

The speaker visibility did not prove to have a significant simple effect on the loudness estimates of speech but significantly interacted with the distance factor.

The loudness estimates depended differently on the speaker distance whether the latter was visible or hidden ($F(4, 64) = 6.535$, $p < .001$). This significant interaction is depicted in Fig. 12, which shows little difference between the two curves at the closest distances. Bonferroni post-hoc tests highlighted that the results significantly differed between the visible and hidden source

when the latter was at 16 m ($p = .007$), with a relatively small effect size ($d = .257$). These loudness-distance functions were fitted to power functions as described in Eq. 1. The fittings revealed $b = -0.05$ when the speaker was visible and $b = -0.08$ when he was hidden.

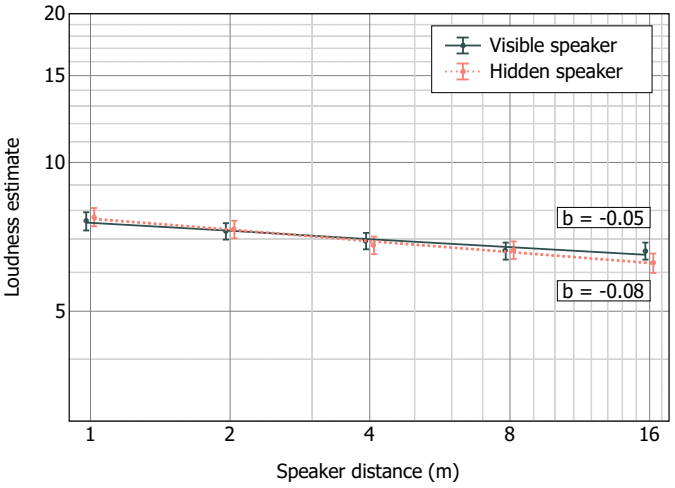

Fig. 12 Geometric mean loudness estimate made when listeners were instructed to focus on the distal stimuli, as a function of the speaker distance made when the speaker was visible (solid line) or hidden (dashed line), with 95% confidence intervals. The b exponent next to each curve depicts the steepness of the slope.

4.3.3 Distance × Level × Visibility

The loudness-distance functions depended on the speaker visibility differently depending on the word production level ($F(4.704^1, 75.258) = 3.015$, $p = .017$). This significant interaction is depicted in Fig. 13, which shows that at low level only, the loudness-distance function obtained when the speaker was hidden is steeper than the one obtained when the speaker was visible. The b exponents depicting the steepness of each curve are presented in Table 8. Each loudness-distance function satisfies $|b| < 0.1$ and the loudness estimates could thus be considered constant in every condition according to the literature.

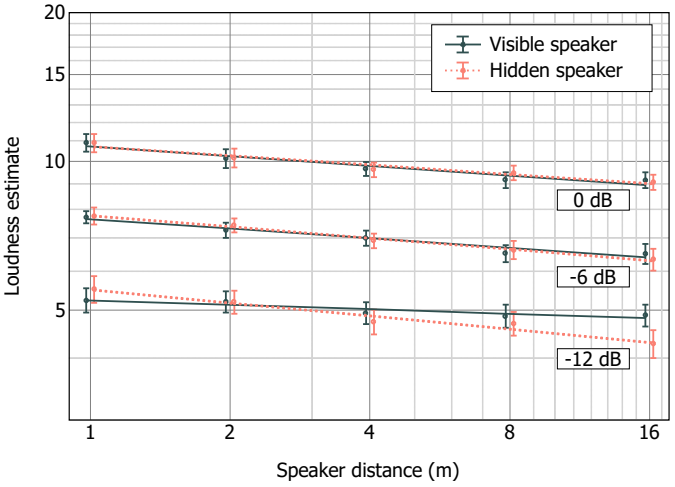

Fig. 13 Geometric mean loudness estimate made when listeners were instructed to focus on the distal stimuli, as a function of the speaker distance made when the speaker was visible (solid line) or hidden (dashed line) for each word production level (0 dB, −6 dB and −12 dB), with 95% confidence intervals. The b exponents depicting the steepness of each slope are given in Table 8.

	b	
Level	Visible	Hidden
0 dB	−0.07	−0.06
−6 dB	−0.07	−0.08
−12 dB	−0.03	−0.09

Table 8 b exponents of the power functions $L = k \cdot r^b$ resulting from the fitting of loudness-distance functions obtained at each word production level and for each visibility condition of the speaker.

4.3.4 Discussion

According to the significant effects involving the distance factor revealed by the ANOVA, the loudness estimates of speech cannot be considered as strictly constant with the speaker distance. The decrease of these estimates with the distance was however weak even in absence of reverberation or visual cues, as depicted in Fig. 12 considering the non-significance of the three way E × D × V interaction (see Table 7), or as depicted in Fig. 13 considering the non-significance of the four way E × D × L × V interaction (see Table 7). The b exponents revealed by the power fittings are low enough to be considered

as describing loudness constancy according to the literature [6]. This can be accounted for by the perceived speaker's vocal effort, which can be deducted from the timbral cues conveyed by speech. Such cues thus provided the listeners with power cues allowing loudness constancy whatever the environment or visibility conditions in which the sounds were presented. Such loudness constancy was achieved whatever the vocal effort with which the words were pronounced, as depicted in Fig. 13, along with the corresponding b exponents given in Table 8.

The extrinsic power and distance information provided by the room acoustics did not prove to reinforce loudness constancy, as no significant interaction involving both the environment and distance factors was revealed by the ANOVA. A strong weight was therefore given to the timbral cues in these loudness evaluation processes, accounting for the fact that this analysis revealed few significant interactions.

5 General discussion and conclusion

The results of these three experiments show that:

When listeners were asked to focus on the proximal stimuli:

- The loudness of noises followed the at-ear sound pressure level and was thus dependent on the sound source distance. It did not depend on the source visibility and thus was not dependent on the perceived source distance (which depended on the visibility).
- The loudness of speech signals depended on the at-ear sound pressure level (and thus on the source distance) in a similar way as the loudness of white noises. The timbre and the perceived vocal effort did therefore not have an effect on the loudness.

When listeners were asked to focus on the distal stimuli:

- The loudness of noises remained constant with the source distance (and thus with the at-ear sound pressure level) providing distance or power cues were available. These cues could be auditory (the direct-to-reverberant energy ratio as a distance cue, the reverberant energy as a power cue) or visual. Auditory and visual cues led to loudness-distance functions presenting similar degrees of constancy.
- The loudness of speech signals was relatively constant with the source distance, even when the sounds were displayed in an anechoic environment without visual cues to their source. In these conditions, participants based their estimates mainly on the speaker's vocal effort, which was perceived through the voice timbre.

These results notably highlight the significance of the instructions given to the participants in a loudness assessment task. When the instructions explicitly request to focus on the proximal stimuli, the listener evaluates loudness in a similar way for noises or speech, despite the timbral cues provided by the latter. When the instructions explicitly request to focus on the distal stimuli, the process involves the perception of an auditory object [14] that relates to a source. This might be irrelevant for non-familiar stimuli displayed in an environment that does not provide information about the source. As an example, for white noise bursts displayed by an invisible source in an anechoic environment, the listener is only able to evaluate loudness on the basis of the at-ear level even when instructed to focus on the distal stimuli. In these conditions, loudness estimates performed when asked to focus on the distal stimuli are similar to those performed when asked to focus on the proximal stimuli, or when no specific request towards focus is made [22]. When provided with sufficient information about the sound source, the listener has the ability to focus

on the distal stimuli and then to report constant loudness evaluation with distance if the sound source remains of constant power. Such information can be provided through auditory or visual cues. The results of the second experiment of the present study show that the information conveyed by the reverberant field (about the source power [6] and distance [23]) and the visibility of the source (about the source distance [44]) led to loudness constancy in a similar way, despite providing information on different aspects of the source. As shown by the third experiment, the timbral cues conveyed by speech enable loudness constancy with distance even in absence of other information about the source. These results tend to show that when asked to focus on the distal stimuli, listeners do not solely take into account the sounds that reach their ears in order to evaluate loudness. Instead, they evaluate loudness by combining the information about the auditory object they have access to, even if no direct information to the source power is provided (e.g. by combining the at-ear sound level and the visually-determined distance of a source displaying a white noise in free field).

The dichotomy between the perception led by the proximal stimulus and the perception led by the distal stimulus has been studied in several domains of perception. Most of this work has been carried out in visual perception, with studies focusing on size [3], shape [2, 4] or color [53] constancy. Norman [54] reviewed papers where this contrast has been reported as "sensory vs. cognitive"; "proximal vs. algorithmic"; or "direct vs. indirect" theories. These theories refer to a theory which assume the content of the proximal stimulus as the determiner of what is perceived vs. a theory which assume interpretative mechanisms in the perceiver, emphasizing the *equivocality* of the stimulus, respectively. In visual perception, the equivocality of the stimulus can refer to e.g. the distance-dependent size of the image that an object of constant

size projects on the retina, or the orientation-dependent shape of the image formed on the retina by a object of constant shape. Such equivocality could similarly refer to the distance-dependent at-ear sound pressure level of the sounds displayed by a source of constant power. Norman [54] asked observers to evaluate which of two objects was the largest (thus, participants were asked to focus on the distal stimuli), with the two objects being of various sizes and placed at various distances from the observer. On the one hand, reaction times were measured and showed that the more similar the two objects were in size, the longer observers were to pinpoint the largest object. On the other hand, the author observed that the extent to which the results depended on the objects distance depended on the difficulty of the task (that is on the difference between the two objects sizes). The author proposed a continuum with direct size perception at one end and indirect size perception at the other rather than a dichotomy between direct or indirect theories. For obvious comparisons, the author assumed that observers based their judgment on direct size perception (that is the size of the image on the retina, or the size of the proximal stimulus, with no additional processing of stimulus information) and for more ambiguous comparisons, the observers used an internalized representation of the stimuli, which relies on a cognitive approach involving an interpretation of the distance cues in order to perceive the size of the distal stimulus. The three experiments presented in the current paper highlight the ability for listeners to process the stimulus information in order to evaluate the loudness of the distal stimulus. Such processing might involve different cognitive processes depending on the stimulus itself. For determining the acoustic power of the sound source, one could e.g. combine the visually-determined distance and the at-ear level of a noise burst displayed in free field by a visible loudspeaker; or interpret the timbre of a hidden speaker pronouncing words in free field by comparison with

the past knowledge of the human voice. However, the proposal of Norman [54] regarding a continuum with direct size perception at one end and indirect size perception at the other could apply to loudness, with translating direct size perception into the loudness as reported when focusing on the proximal stimulus and translating indirect size perception into the loudness as reported when focusing on the distal stimulus.

If a listener is asked to evaluate "whether a sound is loud" without additional specification, several interpretations of the instruction might coexist. Pollack [52] established a comparison between his observations on the loudness of spoken voice and observations made on "size constancy" in visual perception research. If an observer is asked to estimate the size of an object, without additional specification (which could be assimilated to the loudness of a sound in auditory research), it is likely that this observer would not solely estimate the size of the image formed by this object on their retina (which could be assimilated to the loudness evaluated when focusing on the proximal stimulus), but would also take into account the "real" size of this object, as he perceives it (which could be assimilated to the loudness evaluated when focusing on the distal stimulus). On the other hand, if the same observer is specifically asked to evaluate the size of the image formed by this object on their retina (the proximal stimulus), this evaluation might be independent from the perception of the real size of the object.

Loudness is often studied without specifically leading the participant's attention on the proximal or distal stimulus (e.g. "How loud is this sound?"). In this experimental paradigm, the participant is free to interpret the instructions. If the sound is a distant shout, one could perceive it as either quiet (because it is distant, hence the at-ear level is low) or loud (because it is identified as a shout, which means it is displayed loudly). The loudness evaluated

this way might tend toward either the assessment obtained by instructing to focus on the proximal stimulus or the assessment obtained by instructing to focus on the distal one. Inter-individual differences could arise from listeners having different interpretations of the instructions. As a result, the instructions provided to participants in loudness studies should be carefully chosen and reported.

Open practices statement

None of the data or materials for the experiments reported here is available, and none of the experiments was preregistered.

References

[1] Garrigan, P., Kellman, P.J.: Perceptual learning depends on perceptual constancy. Proceedings of the National Academy of Sciences **105**(6), 2248–2253 (2008). https://doi.org/10.1073/pnas.0711878105

[2] Borresen, C.R., Lichte, W.H.: Shape-constancy: Dependence upon stimulus familiarity. Journal of Experimental Psychology **63**(1), 91–97 (1962). https://doi.org/10.1037/h0042085

[3] Kilpatrick, F.P., Ittelson, W.H.: The size-distance invariance hypothesis. Psychological Review **60**(4), 223–231 (1953). https://doi.org/10.1037/h0060882

[4] Epstein, W., Park, J.N.: Shape constancy: Functional relationships and theoretical formulations. Psychological Bulletin **60**(3), 265–288 (1963). https://doi.org/10.1037/h0040875

[5] Shigenaga, S.: The constancy of loudness and of acoustic distance. Bulletin

of the Faculty of Literature of Kyushu University **9**, 289–333 (1965)

[6] Zahorik, P., Wightman, F.L.: Loudness constancy with varying sound source distance. Nature Neuroscience **4**(1), 78–83 (2001). https://doi.org/10.1038/82931

[7] Altmann, C.F., Matsuhashi, M., Votinov, M., Goto, K., Mima, T., Fukuyama, H.: Visual distance cues modulate neuromagnetic auditory N1m responses. Clinical Neurophysiology **123**(11), 2273–2280 (2012). https://doi.org/10.1016/j.clinph.2012.04.004

[8] Epstein, M., Florentine, M.: Binaural loudness summation for speech and tones presented via earphones and loudspeakers. Ear and Hearing **30**(2), 234–237 (2009). https://doi.org/10.1097/AUD.0b013e3181976993

[9] Reynolds, G.S., Stevens, S.S.: Binaural Summation of Loudness. The Journal of the Acoustical Society of America **32**(10), 1337–1344 (1960). https://doi.org/10.1121/1.1907903

[10] Sivonen, V.P., Ellermeier, W.: Directional loudness in an anechoic sound field, head-related transfer functions, and binaural summation. The Journal of the Acoustical Society of America **119**(5), 2965–2980 (2006). https://doi.org/10.1121/1.2184268

[11] Bolles, R.C., Bailey, D.E.: Importance of object recognition in size constancy. Journal of Experimental Psychology **51**(3), 222–225 (1956). https://doi.org/10.1037/h0048080

[12] Mohrmann, K.: Lautheitskonstanz im entfernungswechsel. Zeitschrift fur Psychologie **145**, 145–199 (1939)

[13] Philbeck, J.W., Mershon, D.H.: Knowledge about typical source output influences perceived auditory distance. The Journal of the Acoustical Society of America **111**(5), 1980 (2002). https://doi.org/10.1121/1.1471899

[14] Bizley, J.K., Cohen, Y.E.: The what, where and how of auditory-object perception. Nature Reviews Neuroscience **14**(10), 693–707 (2013). https://doi.org/10.1038/nrn3565

[15] McAdams, S.: Recognition of sound sources and events. In: Thinking in Sound: The Cognitive Psychology of Human Audition. Oxford science publications., pp. 146–198. Clarendon Press/Oxford University Press, New York, NY, US (1993). https://doi.org/10.1093/acprof:oso/9780198522577.003.0006

[16] Berthomieu, G., Koehl, V., Paquier, M.: Does loudness relate to the strength of the sound produced by the source or received by the ears? A review of how focus affects loudness. Frontiers in Psychology **12** (2021). https://doi.org/10.3389/fpsyg.2021.583690

[17] Cusack, R., Decks, J., Aikman, G., Carlyon, R.P.: Effects of location, frequency region, and time course of selective attention on auditory scene analysis. Journal of Experimental Psychology: Human Perception and Performance **30**(4), 643–656 (2004). https://doi.org/10.1037/0096-1523.30.4.643

[18] Kopčo, N., Shinn-Cunningham, B.G.: Effect of stimulus spectrum on distance perception for nearby sources. The Journal of the Acoustical Society of America **130**(3), 1530–1541 (2011). https://doi.org/10.1121/1.3613705

[19] Koehl, V., Paquier, M.: Loudness of low-frequency pure tones lateralized by interaural time differences. The Journal of the Acoustical Society of America **137**(2), 1040–1043 (2015). https://doi.org/10.1121/1.4906262

[20] Culling, J.F., Dare, H.: Binaural loudness constancy. In: van Dijk, P., Başkent, D., Gaudrain, E., de Kleine, E., Wagner, A., Lanting, C. (eds.) Physiology, Psychoacoustics and Cognition in Normal and Impaired Hearing vol. 894, pp. 65–72. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-25474-6_8

[21] Meunier, S., Savel, S., Chatron, J., Rabau, G.: Interindividual differences in directional loudness. The Journal of the Acoustical Society of America **140**(4), 3268–3268 (2016). https://doi.org/10.1121/1.4970368

[22] Stevens, S.S., Guirao, M.: Loudness, reciprocality, and partition scales. The Journal of the Acoustical Society of America **34**(9B), 1466–1471 (1962). https://doi.org/10.1121/1.1918370

[23] Mershon, D.H., King, L.E.: Intensity and reverberation as factors in the auditory perception of egocentric distance. Perception & Psychophysics **18**(6), 409–415 (1975)

[24] Rosenblum, L.D., Fowler, C.A.: Audiovisual investigation of the loudness-effort effect for speech and nonspeech events. Journal of Experimental Psychology: Human Perception and Performance **17**, 976–985 (1991). https://doi.org/10.1037/0096-1523.17.4.976

[25] Wisniewski, M.G., Mercado, E., Gramann, K., Makeig, S.: Familiarity with speech affects cortical processing of auditory distance cues and increases acuity. PLoS ONE **7**(7), 41025 (2012). https://doi.org/10.1371/

journal.pone.0041025

[26] Epstein, M., Florentine, M.: Binaural loudness summation for speech presented via earphones and loudspeaker with and without visual cues. The Journal of the Acoustical Society of America **131**(5), 3981–3988 (2012). https://doi.org/10.1121/1.3701984

[27] Berthomieu, G., Koehl, V., Paquier, M.: Directional loudness of low-frequency noises actually presented over loudspeakers and virtually presented over headphones. Journal of the Audio Engineering Society **67**(9), 11 (2019). https://doi.org/10.17743/jaes.2019.0018

[28] Brunswik, E.: Loudness constancy with distance variant. In: Perception and the Representative Design of Psychological Experiments, 2nd edn., pp. 70–72. Univ of California Press, (1956)

[29] Calcagno, E.R., Abregú, E.L., Eguía, M.C., Vergara, R.: The role of vision in auditory distance perception. Perception **41**(2), 175–192 (2012). https://doi.org/10.1068/p7153

[30] Renner, R.S., Velichkovsky, B.M., Helmert, J.R.: The perception of egocentric distances in virtual environments - A review. ACM Computing Surveys **46**(2), 1–40 (2013). https://doi.org/10.1145/2543581.2543590

[31] Udesen, J., Piechowiak, T., Gran, F.: The effect of vision on psychoacoustic testing with headphone-based virtual sound. Journal of the Audio Engineering Society **63**(7/8), 552–561 (2015). https://doi.org/10.17743/jaes.2015.0061

[32] ISO 3382–1: Acoustics – measurement of room acoustic parameters – part 1: Performance spaces. Standard, International Organization for

Standardization, Geneva, Switzerland (June 2009)

[33] Cabrera, D., Lee, D., Yadav, M., Martens, W.L.: Decay envelope manipulation of room impulse responses: Techniques for auralization and sonification. In: Proceedings of Acoustics 2011, Gold Coast, Australie, p. 5 (2011)

[34] Salmon, F., Hendrickx, É., Épain, N., Paquier, M.: The influence of vision on perceived differences between sound spaces. Journal of the Audio Engineering Society **68**(7/8), 522–531 (2020). https://doi.org/10.17743/jaes.2020.0046

[35] Kronlachner, M.: Plug-in suite for mastering the production and playback in surround sound and ambisonics. Gold-Awarded Contribution to AES Student Design Competition (2014)

[36] Schörkhuber, C., Zaunschirm, M., Höldrich, R.: Binaural rendering of ambisonic signals via magnitude least squares. In: Proceedings of the DAGA, Munich, Germany, p. 4 (2018)

[37] Bernschütz, B.: A spherical far field HRIR/HRTF compilation of the Neumann KU 100. In: Proceedings of the 40th Italian (AIA) Annual Conference on Acoustics and the 39th German Annual Conference on Acoustics (DAGA) Conference on Acoustics, p. 29 (2013)

[38] IEC 61672: Electroacoustics - Sound level meters (2013)

[39] IEC 61260: Electroacoustics - Octave-band and fractional-octave-band filters (2014)

[40] Marks, L.E., Florentine, M.: Measurement of loudness, Part I: Methods,

problems, and pitfalls. In: Florentine, M., Popper, A.N., Fay, R.R. (eds.) Loudness, pp. 17–56. Springer, New York, NY (2011). https://doi.org/10.1007/978-1-4419-6712-1_2

[41] Rébillat, M., Boutillon, X., Corteel, E., Katz, B.F.G.: Audio, visual, and audio-visual egocentric distance perception by moving subjects in virtual environments. ACM Transactions on Applied Perception **9**(4), 1–17 (2012). https://doi.org/10.1145/2355598.2355602

[42] Likert, R.: A technique for the measurement of attitudes. Archives of Psychology **22 140**, 55–55 (1932)

[43] Zahorik, P., Brungart, D.S., Bronkhorst, A.W.: Auditory distance perception in humans: A summary of past and present research. Acta Acustica united with Acustica **91**(3), 409–420 (2005). Publisher: S. Hirzel Verlag

[44] Cutting, J.E., Vishton, P.M.: Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth. In: Perception of Space and Motion, pp. 69–117. Elsevier, (1995)

[45] Kolarik, A.J., Moore, B.C.J., Zahorik, P., Cirstea, S., Pardhan, S.: Auditory distance perception in humans: a review of cues, development, neuronal bases, and effects of sensory loss. Attention, Perception, & Psychophysics **78**(2), 373–395 (2016). https://doi.org/10.3758/s13414-015-1015-1

[46] Da Silva, J.A.: Scales for perceived egocentric distance in a large open field: Comparison of three psychophysical methods. The American Journal of Psychology **98**(1), 119 (1985). https://doi.org/10.2307/1422771

[47] Loomis, J.M., Klatzky, R.L., Philbeck, J.W., Golledge, R.G.: Assessing auditory distance perception using perceptually directed action. Perception & Psychophysics **60**(6), 966–980 (1998). https://doi.org/10.3758/BF03211932

[48] Petersen, J.: Estimation of loudness and apparent distance of pure tones in a free field. Acta Acustica united with Acustica **70**, 5 (1990)

[49] Allen, G.D.: Acoustic level and vocal effort as cues for the loudness of speech. The Journal of the Acoustical Society of America **49**(6B), 1831–1841 (1971). https://doi.org/10.1121/1.1912588

[50] Warren, R.M.: Anomalous loudness function for speech. The Journal of the Acoustical Society of America **54**(2), 390–396 (1973). https://doi.org/10.1121/1.1913590

[51] Fournier, J.-E.: Audiométrie Vocale : les Épreuves D'intelligibilité et Leurs Applications Au Diagnostic, À L'expertise et À la Correction Prothétique des surdités. Maloine, Paris, France (1951)

[52] Pollack, I.: On the measurement of the loudness of speech. The Journal of the Acoustical Society of America **24**(3), 323–324 (1952). https://doi.org/10.1121/1.1906900

[53] Foster, D.H.: Color constancy. Vision Research **51**(7), 674–700 (2011). https://doi.org/10.1016/j.visres.2010.09.006

[54] Norman, J.: Direct and indirect perception of size. Perception & Psychophysics **28**(4), 306–314 (1980). https://doi.org/10.3758/BF03204389