

COVER PAGE

Document downloaded by @DAEL

Thu Nov 26 19:08:55 2026
For personal use

When automatic English translation is provided, only the original document is authentic.
The EAA cannot be held responsible of any translation error

Bibliographical reference

Room-Induced Cues for the Perception of Virtual Auditory Distance with Stimuli Equalized in Level, Arnaud Bidart and Mathieu Lavandier, Acta Acustica **vol. 102** (Number 1), 2016, pp. 159-169

DOI
https://doi.org/10.3813/AAA.918933

Acta Acustica united with Acustica
Vol. 102 (2016) 159 – 169

DOI 10.3813/AAA.918933

Room-Induced Cues for the Perception of Virtual Auditory Distance with Stimuli Equalized in Level

Arnaud Bidart, Mathieu Lavandier
Univ Lyon, ENTPE, Laboratoire Génie Civil et Bâtiment, Rue M. Audin, 69518 Vaulx-en-Velin Cedex, France.
mathieu.lavandier@entpe.fr

Summary
Auditory distance was judged by listeners wearing headphones for a frontal source in a virtual room. Non-individualized head-related transfer functions were used to simulate distances ranging from 0.5 to 10 m and the broadband level was held constant. A single sample of pink noise was used in experiment 1 and a single speech sentence in experiment 2. In experiment 1, some noise waveforms were "cut" to eliminate onsets and offsets, so that the reverberant tails on transients were not accessible to the listeners in these stimuli. Finally, in both experiments, both dichotic (normal binaural) and diotic (interaurally identical) waveforms were used, intermixing trials with dichotic, left-ear diotic, and right-ear diotic stimuli. Results showed differences in distance estimates that were consistent with the use of room coloration as a monaural spectral cue. The cut noises were perceived as closer than those with normal onsets and offsets, suggesting that the reverberant tail on transients could constitute a complementary monaural temporal cue. Dichotic stimuli were perceived as closer than diotic stimuli (only at the longest simulated distance for the noise), consistent with binaural hearing making the signals sound less reverberant.

PACS no. 43.66.Qp, 43.66.Ba, 43.55.Hy, 43.66.Pn

1. Introduction

The apparent distance of a sound source has been shown to rely on multiple acoustic cues [1, 2]. The sound level at the ears decreases with source-listener distance, especially outdoors, so sound level is often a primary cue for distance perception even though its usefulness can depend on a priori knowledge of the stimulus. The sound is filtered during its propagation through the air, which is more absorptive at high frequencies, so a sound traveling farther will contain less energy at high frequencies. The resulting spectrum at the ears is significantly altered for long propagation distances (typically above 15 m), so that it becomes a distance cue. For nearby lateral sources (within a 1-m range from the listener), interaural level differences (ILDs) vary systematically with distance and constitute another cue for its perception [3]. In rooms, the listener receives the direct sound coming from the source and the sound reflections from room boundaries. The direct-to-reverberant energy (D/R) ratio at the ears decreases with increasing source distance, because the direct energy decreases while the reverberant energy remains approximately constant. The D/R ratio is an important cue for distance perception in reverberant environments. However, it is not completely clear "how it is coded in the auditory system. Evidence that the cue is derived from the monaural time-domain signal is contradicted by other data suggesting that it relies on binaural information" [4]. A recent study [5] showed that distance discrimination of nearby lateral sources based on a combination of the D/R and ILDs cues activated an area in the posterior non-primary auditory cortex also associated with the perception of sound direction changes and movement. This result could support the hypothesis that, like the perception of direction, the D/R cue is based on binaural information; however, it was obtained only with lateral sources and when the cue based on ILDs was potentially involved. The aim of the present study was to investigate how listeners use reverberation to evaluate differences in distance in a room.

The D/R ratio[1] has been used to model distance perception [6]. It can be calculated from the room impulse response (direct and reverberant sounds can be separated by windowing this response in the time domain), but listeners do not have direct access to the room response when they listen to sounds which are not impulses. For example, when listening to a continuous noise source, it is not obvious how listeners could extract the direct and reverberant sound levels to compute the D/R ratio [7, 8, 9]. They might be using other co-varying cues such as [10]: rever-

Received 08 July 2015,
accepted 13 October 2015.

[1] The model proposed in [6] is based on an early/late energy ratio rather than D/R ratio, thus using a temporal integration window to separate the sound into early/late components rather than direct/reverberant components.

© S. Hirzel Verlag · EAA

beration tail on transients (temporal envelope), room coloration and/or interaural coherence. As distance increases and D/R ratio decreases, reverberant tails on transients become more audible and can also fill in gaps in amplitude-modulated signals (e.g., reducing speech modulations), possibly providing a distance cue.

The modification of the spectrum associated with room coloration results from two simultaneous effects [11, 12]: the frequency-dependent absorption by room materials and the comb-filtering due to constructive/destructive interference of sound reflections. The comb-filtering – often also called coloration – results in a spectral ripple which is highly dependent on the particular listener and source positions within the room and does not necessarily co-vary with source-listener distance. On the other hand, absorption generally increases with increasing frequency, so the reverberant sound contains less energy at high frequencies than the direct sound. As a result, when distance increases and D/R ratio decreases, the overall sound is low-pass filtered. This spectral change could be a cue for distance perception [1]. It is similar to the air-absorption cue mentioned previously (but is not restricted to long distances). In the present study, room coloration will refer to both absorption and comb-filtering, because they occurred simultaneously and were not manipulated independently. Note, however, that absorption was the hypothesized cue for distance perception.

The interaural coherence is an index of the similarity of a sound across the ears. It is calculated by taking the maximum value of the interaural cross-correlation function. The multiple sound reflections reaching the listener in a room reduce the interaural coherence because the reflections are not the same at the two ears (unless the configuration is symmetrical within the room [13]). The coherence depends on the particular listener and source positions within the room, but the general trend is that coherence decreases with increasing incoherent reverberant energy, so coherence (like D/R ratio) decreases with increasing distance [14] and constitutes another potential cue for its perception.

Previous studies have shown that distance perception at fixed D/R ratios can depend on the source type and direction [15], on the laterality of room reflections [16], and on the similarity of the signals at the two ears [17]. Bronkhorst [16] found that estimated distance in a virtual room increased as a function of the number of reflecting surfaces and depended in particular on the number of lateral reflections. However, it is not clear if the D/R ratio was kept constant while varying the number of reflecting surfaces, so possible variation in D/R ratio complicates the interpretation of the data. Bronkhorst [17] measured the apparent distance of a source while varying interaural coherence. For all tested source distances, distance estimates were reduced when coherence increased, until finally there was no difference in distance estimates across tested distances for diotic signals, suggesting that distance perception depends on binaural cues. Other studies did not find any difference between monaural and binaural listening for apparent distance evaluation [18] or D/R ratio discrimination [10].

The present study investigated which of three potential room-induced cues – reverberation tail on transients, coloration and/or interaural coherence – are used to evaluate distance. All these cues generally co-vary with D/R ratio but they provide different types of information. To evaluate distance based on changes in D/R ratio, listeners could be detecting changes in spectrum (coloration), temporal envelope (reverberation tail) and/or interaural differences (interaural coherence).

Listeners judged auditory distance for a source simulated in a virtual room using headphones. The room simulation used non-individualized head-related transfer functions (HRTFs), as in previous distance studies [9, 18]. Zahorik [19] showed that listeners' performance in judging auditory distance was not impaired (in terms of individual estimates or their standard deviations) by using the HRTFs of another listener. To prevent the use of broadband sound level as a distance cue and to focus attention on reverberation cues, even if it led to unnatural listening conditions, sounds were equalized in level as in other studies [2, 3, 20]. After each sound presentation, the listener was asked to report the perceived distance of the source. Distance evaluations were averaged across listeners.

2. Methods

2.1. Stimuli

The stimuli were designed to investigate the three potential room-induced cues separately. Coloration and reverberation tail on transients were investigated independently of interaural coherence by using diotic listening, for which coherence was constant for all tested distances. Coloration and reverberation tail were investigated separately by comparing stimuli with very similar spectrum but with or without onset and reverberant tail.

The stimuli were obtained by convolving the source signals with binaural room impulse responses (BRIRs) simulated for a frontal source at different distances from a listener in a virtual room. The BRIRs were generated with the software CATT Acoustic, taking into account air absorption. The software database was used to simulate both the sound source and the listener. The source was simulated from anechoic impulse responses measured every 10° around a loudspeaker. Spatialization was obtained using non-individualized HRTFs measured with a dummy head every 5.6° in the horizontal plane and every 15° in the vertical plane. The simulated room was 12-m long, 5-m wide and 2.8-m high (Figure 1). The materials covering its surfaces are described in Table I. The same diffusion coefficient of 0.2 in each frequency band was used for all surfaces. The listener and loudspeaker were simulated on an axis at 14° from the 12-m wall. The ears of the seated listener were placed on either side of a fixed point located 1.3 m from the floor, 1 m from the 5-m wall and 1.5 m from the 12-m wall. The loudspeaker was also 1.3 m from

Table I. Materials used in the room simulation.

surface	material	absorption coefficient (%) by octave (Hz)							
		125	250	500	1000	2000	4000	8000	16000
floor	wood parquet on concrete	4	4	7	6	6	7	8	9
walls	mineral wool on concrete	3	28	69	85	87	86	85	84
ceiling	plasterboard with air-gap	20	15	10	8	4	2	0.1	0.1

Table II. Interaural coherence, reverberation time and D/R ratio at the left (L) and right (R) ears for the source simulated at the six tested distances.

Simulated distance (m)	0.5	1.5	3	4.5	7.5	10
Interaural coherence	1	0.98	0.95	0.85	0.81	0.64
Reverberation time L (s)	1.44	1.53	1.52	1.53	1.51	1.49
Reverberation time R (s)	1.43	1.54	1.52	1.53	1.51	1.49
D/R ratio L (dB)	17.2	9.9	3.8	0.4	-3.8	-6.8
D/R ratio R (dB)	17.1	9.6	4.0	0.5	-4.0	-7.3

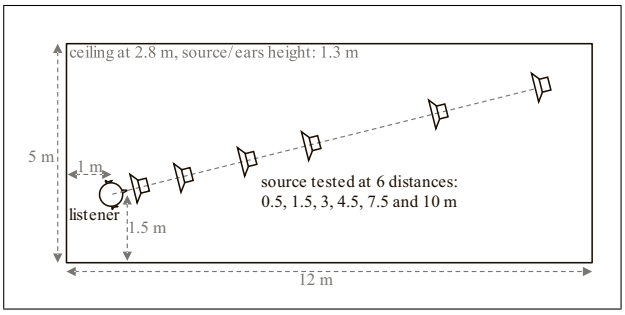

Figure 1. Room simulation layout indicating the listening and source positions.

the floor at six distances in front of the listener: 0.5, 1.5, 3, 4.5, 7.5 and 10 m. These values were chosen because they led to a monotonic decrease of interaural coherence with source distance. Table II presents the broadband interaural coherence, the reverberation time (computed using Schroeder's backward integration method [21]) and the D/R ratio at each ear, obtained for the source simulated at the six tested distances. The BRIRs were digitized at 44.1 kHz with 16-bit quantization.

The source signals were a burst of pink noise and speech. Pink noise is a classic stimulus in distance perception studies. Speech is a very familiar stimulus for listeners, so that they might have developed particular strategies and acquired training to judge distance with this type of signal. The anechoic waveforms (noise and speech) were kept the same in all convolutions. The burst of pink noise was 1-s long, with 10-ms linear onset and offset. The speech signal was a 0.9-s anechoic recording of three words ("Toute la nuit" meaning "All night long") spoken by a male.

In addition, an "altered" stimulus was used: "cut" pink noise. The cut noise was obtained from the pink noise convolved with the six BRIRs, by eliminating the beginning and end of the convolved noises, so that the stationary part of the signals was kept with both onset and reverberant tail eliminated. Each cut noise consisted of the 800 ms of signal taken 100 ms after the beginning of the corresponding convolved pink noise, with 10-ms linear onset and offset. Removing the onset and reverberant tail of the noises had very little effect on their spectrum, so the differences in distance estimates between complete and cut noises could not be accounted for by a change in spectral coloration. On the other hand, for pink noise, the reverberant tails on transients and the associated changes in temporal envelope are only accessible to the listener at the end of the stimuli. The changes in temporal envelope due to reverberation tail are masked in the stationary part of the signals, so they were not accessible to the listener in the cut noises. Comparison of complete and cut noises allowed the investigation of a potential contribution of reverberation tail on transients to distance perception.

To reduce the use of broadband sound level[2] as a distance cue and focus on reverberation cues, all stimuli of a given type were equalized in level [2, 3, 20]. The long-term root-mean-square (RMS) powers of the left and right channels of the stimuli were equalized independently. This broadband level equalization eliminated potential broadband ILDs, but it did not affect the D/R ratio and interaural coherence, nor did it fully compensate for the modifications of source spectra within each frequency band associated with room and head coloration and for the potential associated within-band ILDs. To achieve a comfortable listening level, pink noises were presented at 73 dB SPL (at each ear) and speech at 76 dB SPL. These two signal types were used in separate experiments.

Figure 2 presents the third-octave power spectra of the complete signals simulated at the six tested distances. Spectra are plotted for the left-ear stimuli. The same trends

[2] Equalizing broadband sound level did not prevent loudness from being used as a cue. However, model estimations of loudness level decreased by only 4.5 phons for the noises and speech between 0.5 and 10 m (not shown here). Because sound level was equalized, the variations of estimated loudness were presumably due to the spectral changes associated with room coloration (Figure 2), so that they would have corresponded to the same distance cue.

were observed for the right ear and for the cut noises. These spectra illustrate how room coloration and air absorption systematically decreased the ratio of high- to low-frequency levels as simulated distance increased. Levels above 1250 Hz decreased for both pink noise and speech. It should be noted that the broadband level equalization reduced the level differences below 400 Hz, which is the most energetic part of the spectrum (mainly determining the broadband level).

Three listening modes were used: diotic L and R listening[3], for which the signal arriving at one ear (L for left or R for right) was sent to both headphones, and dichotic listening, for which the left/right ear signal was sent to the left/right headphone, respectively. For dichotic signals, interaural coherence decreased with increasing distance. For diotic signals, coherence was constant and maximum (equal to 1) whereas coloration and reverberant tails on transients were very similar to those of the dichotic signals. Comparison of results for dichotic and diotic L/R listening made it possible to assess the contribution of coherence to distance perception.

In experiment 1, we used pink and cut pink noises in three listening modes (dichotic, diotic L, diotic R). In experiment 2 and in the control experiment, we used speech in these three listening modes.

2.2. Procedure

Digital/analog conversion and amplification of the stimuli were accomplished using a Lynx TWO sound card. Stimuli were presented in randomized order via Sennheiser HD 650 headphones in a double-walled soundproof booth. The response of the headphones was not compensated for[4]. Within experiments 1 and 2, stimuli were presented three times and distance evaluations were averaged across repetitions (arithmetic mean). A computer screen, visible to the listener, was placed outside the booth. The computer keyboard and mouse were inside the booth.

Distance evaluation was based on the approach proposed by Zahorik [15, 22]: at the beginning of the experiment, the listeners were instructed that before each stimulus presentation they should close their eyes and imagine themselves seated in a room in front of a sound source that could vary in distance. The stimulus could be played as many times as desired by pressing any key of the keyboard

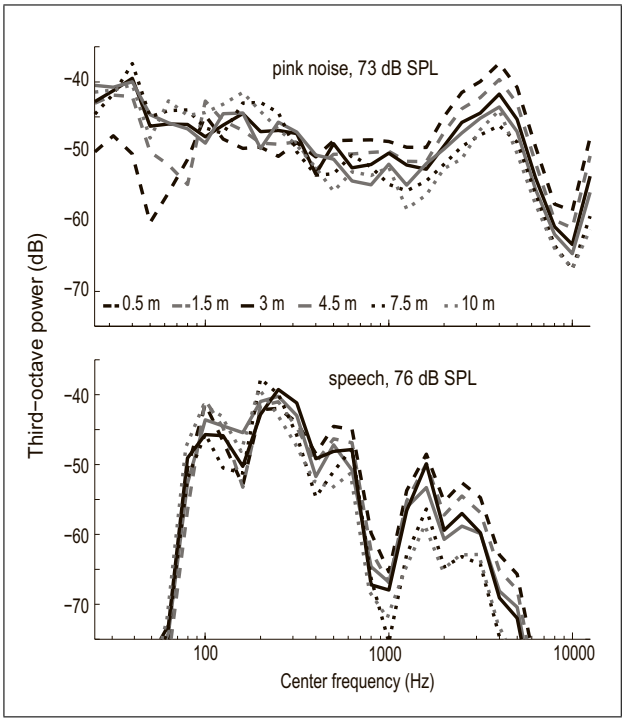

Figure 2. Third-octave power spectra at the left ear for the dichotic complete stimuli (pink noise, speech) simulated at the six tested distances (0.5, 1.5, 3, 4.5, 7.5 and 10 m). The same trends were observed at the right ear and for the cut noises. The sound level at each ear is indicated for each signal type.

(eyes closed). Listeners were asked to report as spontaneously as possible the perceived distance of the sound source. In the present study (experiments 1 and 2), this was done (eyes opened) by moving a cursor on a horizontal continuous linear scale graduated and labelled with distance values every 1 m between 0 and 15 m. Listeners were instructed that 0 m (the default cursor position) should correspond to a sound located inside the head. The next stimulus could be played after they had validated their judgment with the computer mouse. This validation was possible only if the previous stimulus had been listened to at least once and the cursor had been moved.

To check for the potential influence of the restricted range of the scale used to gather distance estimates, a control experiment was run with the stimuli of experiment 2 (speech at six distances in three listening modes). Listeners reported their distance estimates by directly providing a number corresponding to the distance they perceived (in meters) using the keyboard, without any predefined scale. In this experiment, stimuli were presented six times instead of three. Apart from these two differences, the control experiment was identical to experiment 2.

No information concerning the room and distances tested was given. Each experiment started with a practice session to allow listeners to become familiar with the sounds, distances, task and scale. It was short (1 to 2 minutes) and without feedback. Only dichotic stimuli were used for this session. They were played in randomized order only once.

[3] The diotic L and R stimuli were tested separately because they did not contain exactly the same information. If only the R (for example) mode had been tested and a similar effect as the one reported by Bronkhorst [17] had been observed (no difference in distance estimates across tested distances for diotic signals), then it would not have been possible to conclude whether the distance cue which would have disappeared for the diotic R stimuli was the interaural differences or some information received at the left ear.

[4] Because the quality of the headphones ensures that they only marginally altered the spectrum of the reproduced stimuli, in the same way for all conditions and listeners, no effect was expected for not compensating for their response, which was equivalent to using headphones with a perfectly flat frequency response and anechoic stimuli (noise and speech) having a slightly different spectrum.

Table III. Outcomes of repeated-measure analysis of variance (ANOVA) for each experiment. The factors were simulated distance (D), noise type (N) and listening mode (M) in experiment 1 (Figure 3); they were simulated distance and listening mode in experiment 2 and in the control experiment (Figure 4). Sum: Sum of squares, Mean: Mean square.

Factor	Sum	df	Mean	F	p
Experiment 1					
D	7094.7	5	1418.9	197.6	<0.0001
N	307.5	1	307.5	12.7	<0.01
M	25.2	2	12.6	4.7	<0.05
D x N	63.5	5	12.7	8.1	<0.0001
D x M	41.7	10	4.2	2.6	<0.01
N x M	15.7	2	7.8	4.1	<0.05
D x N x M	11.2	10	1.1	0.7	n.s.
Experiment 2					
D	3192.5	5	638.5	143.5	<0.0001
M	71.6	2	35.8	16.4	<0.0001
D x M	48.8	10	4.9	4.6	<0.0001
Control experiment					
D	1511.7	5	302.3	27.7	<0.0001
M	44.7	2	22.3	10.1	<0.001
D x M	21.8	10	2.2	2.1	<0.05

2.3. Listeners

Thirty listeners aged between 22 and 37 yrs (mean age of 25 yrs, standard deviation of 4.5 yrs, 22 males) completed experiment 1 in about 20 minutes and experiment 2 a few days later in about 10 minutes. Fifteen different listeners aged between 22 and 32 yrs (mean age of 26 yrs, standard deviation of 4 yrs, 11 males) completed the control experiment in about 20 minutes. All listeners self-reported normal hearing and were paid for their participation.

3. Results

Figure 3 presents the mean[5] estimated distances measured for the pink and cut pink noises. Estimated distance increased with increasing simulated distance for all three listening modes, the increase being less pronounced for the cut noises than for the complete noises. The main effects of simulated distance and noise type were significant (Table III). Their interaction was significant. Tukey pairwise comparisons indicated that, for each noise type, estimated distances were all significantly different from each other except at 7.5/10 m ($p<0.0001$ in all other cases). Cut noises were perceived as significantly closer than complete noises at all simulated distances except 0.5 m ($p<0.01$ in all other cases).

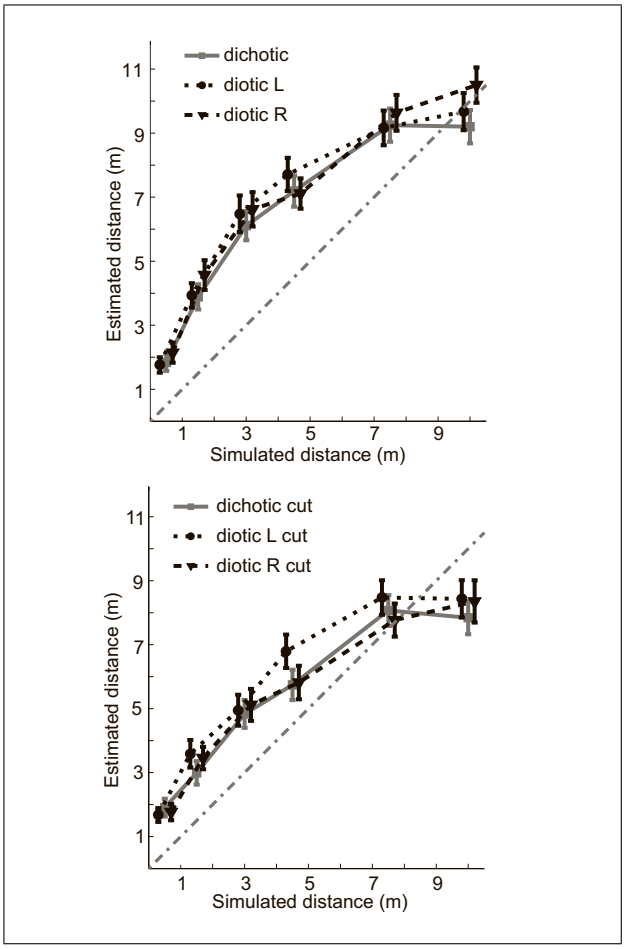

Figure 3. Mean estimated distance with standard error across listeners as a function of simulated source distance for experiment 1, for pink noise (top panel) and cut pink noise (bottom panel) in dichotic (solid grey lines, squares), diotic L (dotted lines, circles) and diotic R (dashed lines, triangles) listening modes. Small horizontal offsets were added to reduce symbol overlap. The dashed grey line indicates perfect performance.

The main effect of listening mode was also significant, and it interacted significantly with each of the effects of simulated distance and noise type. Tukey pairwise comparisons indicated that, for each listening mode, all estimated distances were significantly different from each other, except at 7.5/10 m for all modes and at 3/4.5 m for the diotic R mode ($p<0.01$ in all other cases). For each simulated distance between 0.5 and 7.5 m, the different listening modes did not lead to significant difference in estimated distance. At 10 m, dichotic listening led to estimated distances significantly shorter than those measured for diotic R listening ($p<0.05$). For each listening mode, cut noise was perceived as significantly closer than complete noise ($p<0.001$ in all cases). For each noise type, the different listening modes did not lead to significant differences in estimated distance, except that dichotic listening led to estimated distances significantly shorter than for diotic R lis tening for the complete noise ($p<0.01$).

[5] Individual data were averaged across listeners without being log-transformed. The distribution of the data was normal (Shapiro-Wilk test) for most of the tested conditions: 27 out of 36 in experiment 1 and 17 out of 18 in experiment 2. The number of normally distributed conditions was smaller for log-transformed data (24 in experiment 1 and 13 in experiment 2). The non-normal distributions were due to a few outliers rather than a particular trend in the data, so no other transform was considered. Outliers were not removed from the data. All statistical analyses were done using Statistica 8.0.

Figure 4 presents the mean estimated distances for the speech signals. Estimated distance increased less with increasing simulated distance for dichotic listening than for diotic listening. The main effects of simulated distance and listening mode were significant (Table III). Their interaction was also significant. In experiment 2, for each listening mode, only the simulated distances 3/4.5 m and 7.5/10 m did not lead to significantly different estimated distances ($p<0.0001$ in all other cases). For each short simulated distance (0.5 or 1.5 m), the different listening modes did not lead to significantly different estimated distances. For each longer simulated distance, the two diotic modes did not lead to significantly different estimated distances either, except at 4.5 m ($p=0.04$). For each simulated distance at and above 3 m, dichotic listening led to estimated distances significantly shorter than those measured diotically ($p<0.05$ in all cases), except at 4.5 and 10 m, for which the dichotic and diotic R modes did not lead to significantly different estimated distances.

All the significant differences observed in the control experiment were already observed in experiment 2. Some significant differences of experiment 2 were not observed in the control experiment: the simulated distances 0.5/1.5 m did not lead to significantly different estimated distances for the dichotic and diotic L modes; the diotic modes did not lead to significantly different estimated distances at 4.5 m; the dichotic and diotic L modes did not lead to significantly different estimated distances at 3, 4.5 and 10 m; the dichotic and diotic R modes did not lead to significantly different estimated distances at 7.5 m.

To summarize the results, listeners tended to overestimate source distance between 0.5 and 4.5 m when using the graphical scale to provide their answers (experiments 1 and 2), whereas this overestimation was not observed in the control experiment based on direct number report without predefined scale. Distance was generally underestimated at 10 m using both tasks. Listeners tended to judge cut noises as slightly closer than complete noises (experiment 1), and dichotic speech as slightly closer than diotic speech for simulated distances at and above 3 m (experiment 2).

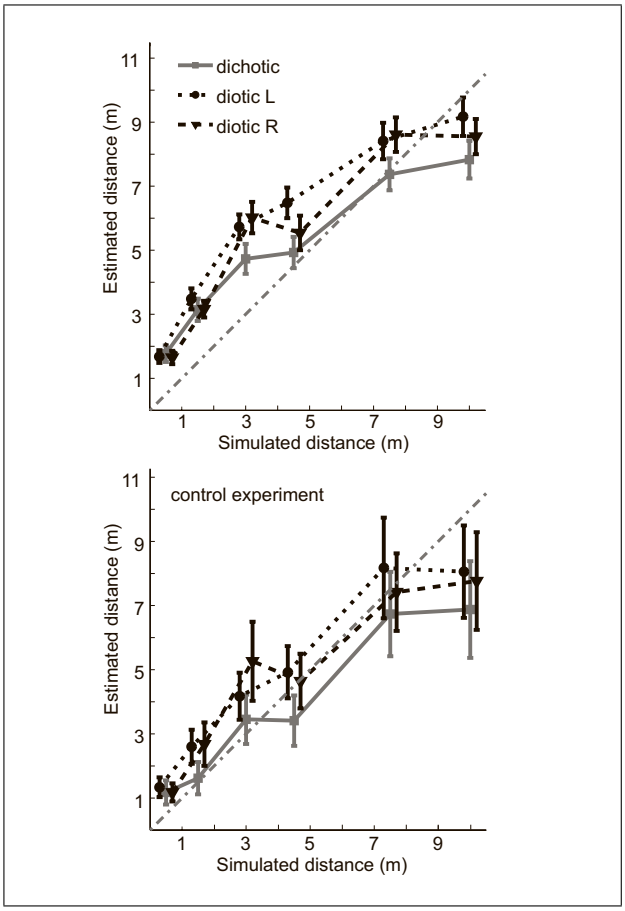

Figure 4. Mean estimated distance with standard error across listeners as a function of simulated source distance for experiment 2 (top panel) and for the control experiment (bottom panel), for speech in dichotic (solid grey lines, squares), diotic L (dotted lines, circles) and diotic R (dashed lines, triangles) listening modes. Small horizontal offsets were added to reduce symbol overlap. The dashed grey line indicates perfect performance.

4. Discussion

4.1. Estimation of simulated distance

The experimental conditions tested in the present study were unnatural. They were investigated to be able to isolate reverberation cues for virtual distance perception and to understand how listeners evaluate distance based on changes in D/R ratio. The broadband sound level of the stimuli was fixed across simulated distances and tested conditions, which would not occur for real-life sources varying in distance. The level was also fixed across ears. This equalization eliminated the broadband ILD, which has been shown to be a distance cue but mainly for lateral sources within a 1-m range from the listener [3]. The equalization did not fully eliminate the within-band ILDs resulting from room and head coloration, which still influenced interaural differences and coherence. The set of tested stimuli was limited to a single sample of pink noise and a single sample of speech. It remains to be tested if listeners would use the same distance cues in more natural conditions, and if so if they would do it in the same way as in the constrained conditions tested here.

Distance was generally underestimated at 10 m. This was probably not due to the limited range of the graphical scale used in experiments 1 and 2, because the same underestimation was observed in the control experiment. If care is taken when defining the upper boundary of the scale (here 15 m for sources simulated at a maximum of 10 m), the restricted range of the scale does not necessarily lead to underestimation at large distances. In experiments 1 and 2, listeners tended to overestimate source distance between 0.5 and 4.5 m. Past work has also shown overestimation, but less than that observed here and limited to relatively close sources (within 1 m from the listener [1, 8]). Distance was not overestimated in the control experiment. This difference indicates that the absolute dis-

tance estimates presented here should be considered with caution. They have been influenced by the reporting task/graphical scale and could also have been influenced by the tested stimuli, in particular the sound level equalization, and the accuracy of the room simulation. On the other hand, the relative distance estimates or trends of variation of estimated distance with simulated distance were very similar in experiment 2 and the control experiment: dichotic speech was generally judged as closer than diotic speech at and above 3 m (even if the differences were rarely significant in the control experiment) and estimated distance seemed to reach an asymptote with increasing simulated distance. This last trend was also observed with the pink and cut pink noises in experiment 1 and is consistent with previous studies highlighting an auditory horizon [2, 6, 15, 20]. Consequently, the results of the present study should be considered in terms of relative differences in estimated distance rather than absolute distance evaluation.

The control experiment involved more repetitions of the stimuli but fewer listeners. It was designed to produce the same number of distance evaluations as experiment 2 but resulted in larger standard errors and reduced statistical power, probably because inter-individual variability was larger than intra-individual variability. Because the trends observed across conditions were very similar in experiment 2 and the control experiment, only the results of experiments 1 and 2 are discussed below, focusing on the cues underlying the observed differences in perceived distance rather than the relation between absolute perceived distance and simulated distance.

4.2. Room coloration

A main effect of simulated distance was observed for speech and all types of pink noise, including diotic cut noise for which the distance cue did not rely on interaural differences and was probably spectral. Room coloration (including both absorption and comb-filtering due to reflections) provided such a cue by systematically decreasing high-frequency levels with increasing distance (because room absorption increases with increasing frequency). The results presented here are thus consistent with the use of room coloration as a monaural spectral cue, as hypothesized by Zahorik *et al.* [1] and in agreement with the results of Larsen *et al.* [10], who found that changes in D/R ratio were discriminated primarily on the basis of room coloration. The present study does not define the exact nature of this coloration cue, which could have been the high-frequency level or the spectral balance (high- to low-frequency level ratio).

Previous studies showed that reducing high-frequency content increased estimated source distance. Coleman [23] measured apparent distance with a click reproduced by loudspeakers in a large room, this click being low-pass filtered at 7680 or 10560 Hz. Estimated distance increased with the attenuation of high-frequency energy. The same effect was measured by Little *et al.* [24] using a fixed loudspeaker reproducing broadband noise low-pass filtered at 5, 6 or 6.7 kHz. Butler *et al.* [18] measured apparent distance using noise burst trains and headphones, the bursts being low-pass (with a cutoff frequency of 2, 1 or 0.5 kHz), high-pass (with a cutoff frequency of 6, 4 or 2 kHz) or broadband noise recorded at a fixed distance in a reverberant or anechoic room. The high-pass noises were perceived as closer than the broadband noise which was perceived as closer than the low-pass noises in both rooms (differences being larger in the reverberant room). Decreasing the cutoff frequencies of low-pass and high-pass noises increased apparent distance. The present study shows that room coloration can provide a distance cue similar to those obtained with artificial filtering in these previous studies.

Kolarik *et al.* [9] measured distance discrimination using headphones for white noise reproduced in a virtual room or anechoic room while using both equalized and original sound levels, so that sound level and reverberation cues could be tested in combination or independently. Participants were either blind or were sighted "controls", who kept their eyes closed while listening. Controls were not able to discriminate white noise distances using only reverberation cues (even though blind participants could). This could be at least partly explained by the level equalization, which could have prevented listeners from using room coloration as a distance cue with white noise. While room coloration produces similar differences in spectral balance across simulated distances for all types of signal, the equalization reduces the level differences in the most energetic parts of the spectrum (which mainly determine the broadband level). For example, in the configuration tested in the present study, levels were at least partly equalized below 400 Hz for pink noise and speech (Figure 2), whereas they would have been equalized above 2 kHz for white noise (not shown here). Thus, if the room coloration cue was the high-frequency level rather than the spectral balance, it was available for pink noise and speech but would not have been available for white noise. This might account for the results of Kolarik *et al.* [9]. The fact that they used relatively short noise bursts (90 ms compared to 800/1000 ms in the present study) could also explain the apparent discrepancy in the results, but differences in stimulus duration between 50 and 600 ms did not influence distance evaluation in [17]. Listeners can perceive the distance of a white noise source when sound level is not equalized [22, 25]. Zahorik [15] even showed that, if realistic sound level variations are preserved, reverberation is used to estimate the distance of a white noise source.

4.3. Reverberation tail on transients and binaural de-reverberation

Figure 5 presents the difference in mean estimated distance between complete and cut noises, measured as a function of simulated source distance in experiment 1. Cut noises were perceived as slightly closer than complete noises, the difference increasing when simulated source distance increased up to 3 m. This difference further in-

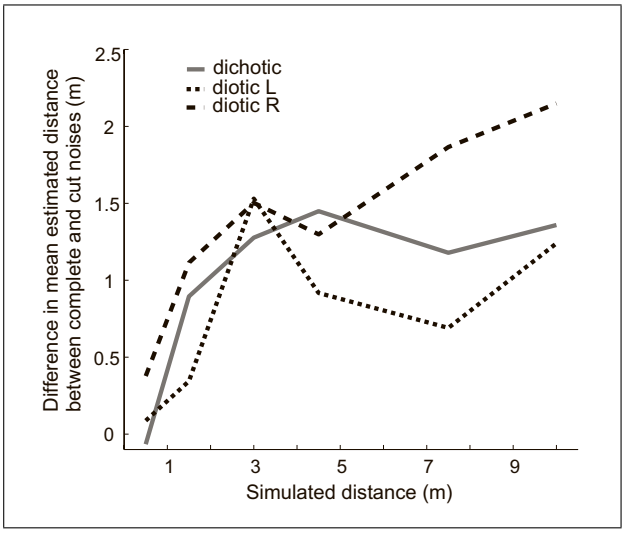

Figure 5. Difference in mean estimated distance between complete and cut noises as a function of simulated source distance for experiment 1 (Figure 3) in dichotic (solid grey line), diotic L (dotted line) and diotic R (dashed line) listening modes.

creased with simulated distance above 3 m for the diotic R mode, whereas it remained approximately constant for the dichotic mode. For the diotic L mode, the difference between complete and cut noises decreased with simulated distance between 3 and 7.5 m and increased again for a source simulated at 10 m. Removing the onset and reverberant tail of noises thus reduced estimated distance, the reduction being generally stronger at larger distances, which correspond to longer and relatively more energetic reverberant tails in the room impulse responses. This effect was also observed for diotic listening, suggesting that the corresponding distance cue was based on monaural information. Complete and cut noises had a slightly different duration, 1 s (plus reverberant tail) and 0.8 s respectively. However, as discussed above, differences in stimulus duration did not influence perceived distance in [17] and the main difference in duration was here constant across simulated distances (discarding the comparatively short duration of the tails), indicating that duration was probably not used as a distance cue in the present study either. Note that onset and reverberant tail were eliminated simultaneously in the cut noises, so their potential individual influences on distance perception cannot be estimated here. Because the differences in onset and reverberant tail of the stimuli across conditions result from the differences in reverberant tail of the room impulse responses, in the following discussion reverberant tail is used for reverberant tail on transients, refering simultaneously to the onsets and offsets in the stimuli.

In experiment 1, the sound level of the reverberation tail at the end of the complete noises increased with increasing simulated distance, so this level could have been an underlying cue for distance perception: the lower this level was, the closer the source was perceived. This sound level could also have been associated with the level of the direct sound, which might be accessible to the listener at the onset of the noise, in order to compute the D/R ratio and use it directly to judge distance (rather than using other covarying cues). Finally, the reverberation tail is also associated with the percept of reverberance, which corresponds to the perceived amount of reverberation (how much it "sounds" reverberant) and depends on the level of the reverberation tail but maybe also on other parameters such as its duration or coherence. Reverberance could also have been an underlying cue for distance perception: the less the source sounded reverberant, the closer it was perceived. The present study cannot tease apart these three possible interpretations concerning the nature of the distance cue based on reverberation tail. Note that the overall level equalization used increased the level of the reverberant tail with increasing distance. The equalization potentially reinforced the reverberation tail cue for distance, if this cue was associated with the absolute rather than relative level of the tail. In experiment 1, the effect of reverberation tail on distance perception was smaller than the spectral effect of room coloration: the largest difference in distance estimate between cut and complete noises (between 1.5 and 2 m, see Figure 5) was considerably smaller than the difference in estimate for cut noise simulated at 0.5 and 10 m (between 6 and 7 m, see Figure 3).

Bronkhorst [17] did not find any effect on distance perception of varying the linear onset and offset of the noise envelope of the source. Also, Zahorik [7] showed that D/R ratio discrimination was similar for source signals differing in temporal structure. However, there is a difference between the signal-at-the-source offset manipulated by these authors and the signal-at-the-ear offset manipulated in this study. Here the sound offset was removed after convolution by the room response (thus removing the reverberant tail at the end of the stimuli), which is a much stronger manipulation than varying the source signal offset and then convolving the signal with the room response (thus keeping the reverberant tail at the end of the stimuli). This might explain why Bronkhorst and Zahorik did not find a significant effect of source signal onset and offset, whereas experiment 1 indicated a significant contribution of the reverberation tail at the end of the noises to the perception of their distance. Larsen *et al.* [10] also found that temporal cues (in particular sound decay) were used to discriminate the D/R ratio, especially when comparatively more potent spectral cues were diminished by band-limiting the signals.

In experiments 1 and 2, dichotic stimuli were perceived as slightly closer than diotic stimuli (this effect being stronger for the speech, limited to the longest simulated distance for the noise, and opposite to the effect observed in [17] when varying noise coherence). Again, reverberation tail on transients and reverberance could have been the underlying cue here. Dichotic stimuli would have been perceived as closer because they sounded less reverberant than diotic stimuli. This interpretation is consistent with the "squelching" effect of binaural hearing described by Koenig [26]: reverberance is perceived as lower when it is heard binaurally rather than monaurally. The binaural sys-

tem partly de-reverberates the signals. For example, binaural hearing was shown to reduce the loss of intelligibility associated with the temporal smearing of speech by reverberation [27, 28, 29]. This has been recently interpreted as spatial release of masking from the room reflections that reach the listener with different azimuths than the direct speech [30]. The binaural system would de-reverberate the signal by unmasking its early part from its late (detrimental/masking) part coming from different directions, in the same way that it can unmask a target signal from a spatially-separated masking source. Binaural hearing could then have reduced perceived distance by reducing the perceived level of the reverberation tail on transients. Butler *et al.* [18] measured the apparent distance of low-pass, high-pass and broadband noise-burst trains in a reverbeant room, using monaural and binaural listening over headphones. There was no main effect of listening mode, but the low-pass noises (judged as farther than the other noises) tended to be judged as closer in binaural listening, as observed with different stimuli in the present study.

The difference in distance estimate between dichotic and diotic listening increased with simulated distance. This is consistent with the fact that reverberation tail on transients becomes longer and relatively more energetic with increasing simulated distance, so the amount of reverberance that can be reduced by the binaural system also increases with distance. In experiment 1, dichotic listening led to shorter estimated distances than diotic R listening for the complete noise but not for the cut noise. The latter had no transient with a reverberant tail and so very limited reverberance to be reduced by binaural hearing. For the same reason, the reduction of perceived distance when listening dichotically was more evident with speech (Figure 4) than with complete pink noise (Figure 3, dichotic and diotic R modes led to different estimated distances only at 10 m), probably because speech contained more transients and unmasked reverberant tails than noise. With amplitude-modulated signals such as speech, reverberant tails also fill in gaps in the modulations and the resulting loss of modulations could constitute a complementary cue for the perception of reverberance and/or distance.

The diotic L and R stimuli were tested separately because they did not contain exactly the same information. The tested configuration was not symmetric (Figure 1), so the pattern of reflections arriving at the ear was not identical for the left (L) and right (R) ears. However, it seems that the corresponding differences only marginally affected distance perception in the present study. The only significant difference in distance estimate between the diotic modes was observed only for speech and at a single simulated distance (4.5 m), for which the diotic R speech was perceived as slightly closer than the diotic L speech (as a result, the dichotic and diotic R modes did not lead to significantly different estimates at this distance). It was also observed that, at 10 m, while the diotic modes did not lead to significantly different estimates, dichotic listening led to estimated distances significantly shorter than those measured for diotic R (but not diotic L) listening with noise and for diotic L (but not diotic R) listening with speech. This dependence of the results on the type of signal considered – while both types were convolved with the same L and R impulse responses – could indicate that measurement variability might be responsible for the fact that the limited effect of binaural de-reverberation did not lead to significant differences in distance estimate between dichotic and diotic listening modes in all tested conditions.

4.4. Interaural differences and coherence

Bronkhorst [17] found that the estimated distance of a pink noise source decreased with increasing interaural coherence, with no difference in perceived distance across tested source distances for diotic signals. In experiment 1, the estimated distance did not decrease when the coherence of the pink noises increased (from dichotic to diotic listening), indicating that coherence was probably not used as a distance cue in the present study. However, the two studies differed in many experimental aspects. The simulated rooms and distances were different. Non-individualized HRTFs were used here whereas Bronkhorst used individualized HRTFs. However, using individualized HRTFs should favor distance perception rather than limit it.

The two studies controlled sound level in different ways: in the present study, stimuli were equalized in overall level, whereas Bronkhorst equalized his stimuli based on the level of the direct sound and roved the overall level during the experiment. Natural broadband level variations with distance were eliminated with both methods and listeners did not use broadband level in their distance evaluations. However, keeping some level variations could lead to a more realistic simulation [15, 17] and make it easier for a listener to imagine real-life sources, because normally level is an important cue for distance perception. On the other hand, roving the overall level could also prevent listeners from using the coloration cue highlighted here, if this cue relies on the high-frequency level rather than the spectral balance. This difference in experimental protocol could at least partly explain the conflicting results between the two studies.

Bronkhorst [17] also made sure that listeners could get used to the acoustics of the tested room. Before and during the experiment, trials were included in which listeners could compare the original binaural stimuli and switch at will between simulated distances. The aim was to ensure that modifications in the stimuli did not disturb the adaptation to the simulated room, and to check that listeners really heard the reference virtual source at the same location as the real source. Even if listeners seemed able to estimate distance in our experiments, nothing was done to preserve/control accommodation to the room.

In the present study, listeners were asked to close their eyes and imagine themselves seated in a (virtual) room in front of a (virtual) sound source. In the experiment of Bronkhorst [17], listeners were seated in the actual room used to produce the stimuli, with eyes open in front

of a (silent) loudspeaker that was used as a visual anchor. Further visual anchors were provided by two poles placed between them and the loudspeaker. One possible reason why the listening mode did not have the same influence in the two studies is that the present study measured virtual auditory distance with eyes closed, whereas Bronkhorst measured auditory distance with eyes open. Zahorik [31] showed that vision allows more accurate and less variable distance estimation, but visible targets can also capture (and potentially bias) the perceived position of a sound source [1]. Measuring auditory distance with eyes open and visual anchors might require perfect sound externalization to avoid any mismatch between the auditory and visual positions. Hartmann and Wittenberg [32] and Catic *et al.* [33] evaluated externalization (eyes opened) by comparing the sound reproduction of a frontal loudspeaker in a room with a simulation of this reproduction using headphones. Reducing the binaural information in the headphone signals reduced externalization in both studies, causing a mismatch between the auditory and visual positions. Catic *et al.* [33] stated that there was no externalization for diotic listening. These results could explain the decrease of estimated distance with increasing interaural coherence (decreasing binaural information) measured in [17]. Listeners might not have been able to evaluate auditory distance, using for example room coloration, if the mismatch between externalization and visual anchors was too great.

Only one participant directly reported an issue with the externalization of the stimuli after participating in the present study. Listeners were instructed to reply 0 m for a sound located inside the head. This participant provided 49 zeros out of 108 distance estimates in the control experiment, another participant provided 29 zeros, four participants provided between 1 and 8 zeros, and nine participants never replied 0 m. In experiment 1 involving 108 estimates, the number of reported zeros was: 18 for one listener, 13 for another listener, between 1 and 6 for nine participants, and nineteen participants never replied 0 m. In experiment 2 involving 54 estimates, the number of reported zeros was: 11 for one listener, between 1 and 3 for five participants, and twenty-four participants never replied 0 m. A very large majority of the zeros were reported for the two shortest simulated distances (0.5 and 1.5 m). Note that listeners were asked to estimate virtual distance, not directly externalization. Even if Zahorik [19] showed that distance estimation was not impaired by using the HRTFs of another listener, externalization should be reduced when using non-individualized HRTFs [9, 18, 19], monaural listening [18] or diotic listening. However, there might also be a difference in the level of externalization required for distance evaluation with eyes closed or eyes open with the necessity to match visual anchors. When there is no visual anchor, there is no mismatch between reduced externalization and vision. Listeners can potentially imagine themselves in front of the source, imagine its distance, and evaluate virtual auditory distance more independently of the degree of externalization. This difference between virtual auditory distance and auditory distance remains an open question to be investigated.

5. Conclusions

Several studies from the literature have shown that reverberation is used to estimate auditory distance. The present study investigated which of several co-varying cues associated with reverberation could be used to evaluate the distance of frontal sources simulated in a virtual room using non-individualized HRTFs and stimuli equalized in sound level. For a sample of pink noise and a single speech sentence, the main differences in distance estimate were consistent with the use of room coloration as a monaural spectral cue. Cut noises without normal onset and offset were perceived as slightly closer than the corresponding complete noises, suggesting that reverberation tail on transients could constitute a complementary temporal cue. This cue was based on monaural information but was affected by binaural hearing: dichotic stimuli were perceived as slightly closer than diotic stimuli (only at the longest simulated distance for the noise), consistent with binaural hearing making the signals sound less reverberant. This effect was more evident for speech which contained more transients and amplitude modulations. It remains to be tested whether these results generalize to more natural conditions, including in particular sound level variations.

Only one room was investigated in the present study. Coloration and reverberation tail are room dependent. For example, rooms with a different absorption/reverberation time can lead to a different reverberation tail at the same source-listener distance and/or to different quantitative variations of the reverberation tail with distance. Thus, the usefulness of coloration and reverberation tail as distance cues probably depends on the room. It should be investigated if and how listeners can make reliable distance estimates across different rooms. Interaural coherence was probably not used as a distance cue in the present study. Our failure to replicate the decrease of estimated distance with increasing coherence measured by Bronkhorst [17] could be due to our experimental protocol not including realistic level variations or room accommodation. It could also suggest a difference between virtual auditory distance measured with eyes closed and auditory distance evaluated with eyes open, the latter requiring more binaural information and externalization.

Acknowledgments

The authors would like to thank Nicolas Grimault, John Culling, Sabine Meunier, Leslie Bernstein, Pavel Zahorik, Adelbert Bronkhorst, Brian Moore, Steve Colburn and anonymous reviewers for useful discussions and comments on this study, Joachim Blanc-Gonnet for programming the measurement software, and all listeners involved in the study. Part of this work was performed within the Labex CeLyA (ANR-10-LABX-0060/ ANR-11-IDEX-0007).

References

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

[2] D. H. Mershon, J. N. Bowers: Absolute and relative cues for the auditory perception of egocentric distance. Perception **8** (1979) 311–322.

[3] D. S. Brungart, N. I. Durlach, W. M. Rabinowitz: Auditory localization of nearby sources. II. Localization of a broadband source. J. Acoust. Soc. Am. **106** (1999) 1956–1968.

[4] A. W. Bronkhorst, P. Zahorik: The direct-to-reverberant ratio as cue for distance perception in rooms (A). J. Acoust. Soc. Am. **111** (2002) 2440.

[5] N. Kopčo, S. Huang, J. W. Belliveau, T. Raij, C. Tengshe, A. J.: Neuronal representations of distance in human auditory cortex. Proceedings of the National Academy of Sciences **109** (2012) 11019–11024.

[6] A. W. Bronkhorst, T. Houtgast: Auditory distance perception in rooms. Nature **397** (1999) 517–520.

[7] P. Zahorik: Direct-to-reverberant energy ratio sensitivity. J. Acoust. Soc. Am. **112** (2002) 2110–2117.

[8] N. Kopčo, B. G. Shinn-Cunningham: Effect of stimulus spectrum on distance perception for nearby sources. J. Acoust. Soc. Am. **130** (2011) 1530–1541.

[9] A. J. Kolarik, S. Cirstea, S. Pardhan: Evidence for enhanced discrimination of virtual auditory distance among blind listeners using level and direct-to-reverberant cues. Exp. Brain Res. **224** (2013) 623–633.

[10] E. Larsen, N. Iyer, C. R. Lansing, A. S. Feng: On the minimum audible difference in direct-to-reverberant energy ratio. J. Acoust. Soc. Am. **124** (2008) 450–461.

[11] J. L. Flanagan, R. C. Lummis: Signal processing to reduce multipath distortion in small rooms. J. Acoust. Soc. Am. **47** (1970) 1475–1481.

[12] B. Collin, M. Lavandier: Binaural speech intelligibility in rooms with variations in spatial location of sources and modulation depth of noise interferers. J. Acoust. Soc. Am. **134** (2013) 1146–1159.

[13] M. Lavandier, J. F. Culling: Prediction of binaural speech intelligibility against noise in rooms. J. Acoust. Soc. Am. **127** (2010) 387–399.

[14] W. M. Hartmann, B. Rakerd, A. Koller: Binaural coherence in rooms. Acta Acustica united with Acustica **91** (2005) 451–462.

[15] P. Zahorik: Assessing auditory distance perception using virtual acoustics. J. Acoust. Soc. Am. **111** (2002) 1832–1846.

[16] A. W. Bronkhorst: Modeling auditory distance perception in rooms. Proceedings of Forum Acusticum 2002, 2002.

[17] A. W. Bronkhorst: Effect of stimulus properties on auditory distance perception in rooms. – In: Physiological and Psychological Bases of Auditory Function. D. J. Breebaart, A. J. M. Houtsma, A. Kolrausch, V. F. Prijs, R. Schoonhoven (eds.). Shaker, Maastricht, The Netherlands, 2001, 184–191.

[18] R. A. Butler, E. T. Levy, W. D. Neff: Apparent distance of sounds recorded in echoic and anechoic chambers. J. Exp. Psychol. Hum. Percept. Perform. **6** (1980) 745–750.

[19] P. Zahorik: Auditory display of sound source distance. – In: Proceedings of the 8th International Conference on Auditory Display, ICAD 2002. R. Nakatsu, H. Kawahara (eds.). Kyoto, Japan, 2002, 239–243.

[20] M. A. Akeroyd, S. Gatehouse, J. Blaschke: The detection of differences in the cues to distance by elderly hearing-impaired listeners. J. Acoust. Soc. Am. **121** (2007) 1077–1089.

[21] M. R. Schroeder: New method of measuring reverberation time. J. Acoust. Soc. Am. **37** (1965) 409–412.

[22] P. Zahorik, F. L. Wightman: Loudness constancy with varying sound source distance. Nat. Neurosci. **4** (2001) 78–83.

[23] P. D. Coleman: Dual role of frequency spectrum in determination of auditory distance (L). J. Acoust. Soc. Am. **44** (1968) 631–632.

[24] A. D. Little, D. H. Mershon, P. H. Cox: Spectral content as a cue to perceived auditory distance. Perception **21** (1992) 405–416.

[25] P. D. Coleman: Failure to localize the source distance of an unfamiliar sound. J. Acoust. Soc. Am. **34** (1962) 345–346.

[26] W. Koenig: Subjective effects in binaural hearing. J. Acoust. Soc. Am. **22** (1950) 61–62.

[27] A. K. Nábělek, P. K. Robinson: Monaural and binaural speech perception in reverberation for listeners of various ages. J. Acoust. Soc. Am. **71** (1982) 1242–1248.

[28] J. P. Moncur, D. Dirks: Binaural and monaural speech intelligibility in reverberation. J. Speech Hear. Res. **10** (1967) 186–195.

[29] M. Lavandier, J. F. Culling: Speech segregation in rooms: Monaural, binaural, and interacting effects of reverberation on target and interferer. J. Acoust. Soc. Am. **123** (2008) 2237–2248.

[30] T. Leclère, M. Lavandier, J. F. Culling: Speech intelligibility prediction in reverberation: towards an integrated model of speech transmission, spatial unmasking and binaural dereverberation. J. Acoust. Soc. Am. **137** (2015) 3335–3345.

[31] P. Zahorik: Estimating sound source distance with and without vision. Optom. Vis. Sci. **78** (2001) 270–275.

[32] W. M. Hartmann, A. Wittenberg: On the externalization of sound images. J. Acoust. Soc. Am. **99** (1996) 3678–3688.

[33] J. Catic, S. Santurette, J. M. Buchholz, F. Gran, T. Dau: The effect of interaural-level-difference fluctuations on the externalization of sound. J. Acoust. Soc. Am. **134** (2013) 1232–1241.