



# Effects of interpersonal familiarity on the auditory distance perception of level-equalized reverberant speech

Özgen Demirkaplan and Hüseyin Hacıhabiboğlu\*

Middle East Technical University (METU), Graduate School of Informatics, 06800 Çankaya, Ankara, Turkey

Received 3 June 2020, Accepted 2 November 2020

**Abstract** – Familiarity with sound sources is known to have a modulatory effect on auditory distance perception. However, the level of familiarity that can affect distance perception is not clearly understood. A subjective experiment that aims to investigate the effects of interpersonal familiarity on auditory distance perception with level-equalized stimuli is reported. The experiment involves a binaural listening task where different source distances between 0.5 and 16 m were emulated by convolving dry speech signals with measured binaural room impulse responses. The experimental paradigm involved level-equalized stimuli comprising speech signals recorded from different-gender couples who have self-reported to have known each other for more than a year with daily interaction. Each subject judged the distances of a total of 15 different speech stimuli from their partner as well as spectrally most similar and most dissimilar strangers, for six different emulated distances. The main finding is that a similar but unfamiliar speaker is localized to be further away than a familiar speaker. Another finding is that the semantic properties of speech can potentially have a modulating effect on auditory distance judgements.

**Keywords:** Auditory distance perception, Spatial hearing, Auditory cognition

## 1 Introduction

The capability to localize sound sources in three dimensions provides distinct evolutionary advantages and is one of the two major ways alongside stereoscopic vision that humans can perceive space. Localization of sound sources in their azimuth and elevation has been extensively studied and the psychophysical mechanisms that govern subjective localization in these dimensions have largely been identified [1]. However, auditory distance perception, despite having exceptional theoretical relevance and practical importance, has not been studied in as much detail. Even less attention had been given to its cognitive aspects. This article presents the results of a study investigating the effects of familiarity on distance perception.

Unlike directional localization which mainly depends on interaural and spectral cues, distance perception occurs as a result of the fusion and interplay of many different relative cues that have physical, perceptual and cognitive origins. While the effects of certain perceptual cues such as intensity, direct-to-reverberant energy (D/R) ratio, sound source spectrum, auditory parallax and vocal effort on auditory distance perception are relatively easy to quantify, they do not constitute the only relevant cues [2]. These cues can be summarized as follows.

Intensity of point-like sources decrease by 6 dB per doubling of distance in acoustic free field, also known as the “inverse-square law”, which acts as a dominant distance cue [3]. However, real sound sources are rarely point-like except at low frequencies. Since distance attenuation may deviate from the inverse-square law for sources that are not point-like, intensity may not always convey reliable distance information either as an absolute or as a relative cue when it is the only distance cue [4].

The ratio of the energies of the early and late parts of sound in a room aids the localization of a sound source. This ratio, also known as the D/R ratio, is known to be another dominant auditory distance cue [5]. Experiments show that the human ability to separately perceive the direct sound in a reverberant field depends on both the time delay between the direct sound and the reverberant components and the D/R ratio [4, 6, 7]. Although listeners are sensitive to changes in D/R ratio while estimating distance in a room, it is not known for certain whether D/R ratio or some other parameter that covaries with the latter is used as a cue [8].

Auditory distance perception is also affected by spectral cues [9]. The most likely explanation for this effect is atmospheric absorption which attenuates higher frequencies more than the lower frequencies [10]. Another acoustic binaural cue related to distance perception is auditory parallax [1] where interaural ear attributes change with the source distance.

\*Corresponding author: [hhuseyin@metu.edu.tr](mailto:hhuseyin@metu.edu.tr)

An acoustic cue which is specifically related to conversational contexts is vocal effort. Distance judgements given to the same speech message whispered, spoken normally and shouted at the same nominal sound pressure level are different [11]. In comparison with normal speech, the distance of whispered speech is underestimated and the distance of shouted speech is overestimated. More specifically, in a communication context, the speaker adjusts their speech to overcome the degrading effect of distance attenuation and reverberation and to maintain the intelligibility of speech by the addressed party [12, 13]. Such articulatory adjustments include changes not only in the sound pressure level, but also in the spectrotemporal properties of speech [14]. Familiarity with levels of vocal effort is known to aid in distance perception [15, 16]. A more recent study showed that in a conversational setting where the speaker kept eye contact with the listener while speaking, visually perceived source distance and acoustic properties of the enclosure had significant effects on power, phonation time ratio, and the mean and standard deviation of fundamental frequency of the produced speech [17]. Such studies show the strong effect of vocal effort and more generally the contextual cues in distance perception in a conversational context. Presence of a visual anchor is also known to support auditory distance perception in interaction with the other acoustic cues [18, 19].

Despite the strong impacts of perceptual cues summarized above, they are not the only cues that humans use to estimate the location of a sound source, and non-perceptual factors can also influence estimates of perceived distance of a sound source [4]. Previous studies have shown that affective state of cognition can be a modulating factor in auditory distance perception. For example, the participants in a fearful state judged the distance to the target to be closer than participants in a neutral state, judging targets to be reachable, at distances that were further than for the neutral group [20]. Familiarity with the sound source can also have a strong effect on auditory distance perception. Wisniewski et. al. compared the distance judgements of listeners to stimuli spoken in their native language (semantically and acoustically familiar) with stimuli spoken in a foreign language (semantically unfamiliar but acoustically familiar) and time-reversed speech (neither semantically nor acoustically familiar). The study showed that familiarity with the speech features could improve the accuracy of distance judgements [21]. Results of another, earlier study indicate that increasing familiarity with a sound source and its surrounding environment results in more accurate distance judgements [22].

Previous studies on auditory distance perception also show that humans tend to overestimate distances of sound sources in the peripersonal space and they tend to underestimate the distances of sound sources in the extrapersonal space [6, 23]. Auditory distance perception in the front direction is mostly accurate for sound sources positioned approximately 1 m away from the listener.

Two representations of distance perception were previously proposed. The first representation relates the direct-to-reverberant energy ratio to subjective distance in rooms

[5], combining strictly physical properties such as the reverberation radius of the room and the directivity factor of the sound source. A second, more parsimonious representation [7] is a compressed power function relating actual to perceived distance in a way similar to Stevens' power law [24]. Both representations incorporate asymptotic behavior corresponding to an *auditory horizon* beyond which no source is localized. Similarly both representations capture the experimental findings that auditory distances of sources closer than about 1 m are overestimated and of those further than about 1 m are underestimated.

This paper is concerned with the effect of interpersonal familiarity on auditory distance perception. Based on this, the hypothesis under test is that, distance of a familiar speaker is localized more accurately than the distances of unfamiliar speakers uttering the same sentences. We investigate this hypothesis in a distance perception experiment where the subjects rate the distances of familiar and unfamiliar speakers in a binaural listening task with level equalized stimuli. We presuppose that individuals who have daily interactions for an extended period of time have familiarity to each other's voices. Therefore, the cohort of participants employed in this study comprised couples who know each other for at least one year and have daily interactions, also making sure that each participant personally knew only their partner and not the other participants. The participants were asked to judge the distances of binaural stimuli emulating different source distances using speech signals consisting of five balanced sentences spoken by the participant's partner as well as spectrally most similar and most dissimilar participants to the latter. Our initial assumption was that an unfamiliar speaker's voice with a high long-term spectral similarity to a familiar speaker's voice would be localized less accurately than the familiar speaker and more accurately than the unfamiliar speaker with a low spectral similarity.

Similar to some earlier auditory distance perception studies [25–29] that aimed to investigate reverberation cues for auditory distance perception and how participants perceive distance based on changes in D/R ratio, this study also employs stimuli equalized for overall broadband level. This was done in order to eliminate the effect of level as the most dominant distance cue leaving D/R ratio as the main distance cue. The main intention herein was to make it easier to isolate a posteriori, the effect of interpersonal familiarity as a modulating factor in distance perception of reverberant speech.

The results indicated that the familiarity and the semantic content of the stimuli as well as their interaction have statistically significant effects on the distance judgements. However, this effect did not occur in the way that we initially predicted. Our findings are: (1) spectrally most similar speaker to a familiar speaker is localized further away than a familiar speaker, and (2) the distances of a familiar speaker and the unfamiliar speaker with the lowest long-term spectral similarity are localized similarly. These findings indicate that while interpersonal familiarity is a modulating factor of distance perception, its effect is non-linear.

## 2 Method

### 2.1 Participants

A total of 24 unpaid participants comprising 12 different-gender couples, ranging in age from 19 to 30, participated in the study. All couples were partners for at least one year with daily interaction. The participants acted in two ways: (1) their speech was recorded to be used as stimuli in the experiment, and (2) they provided distance judgments in the experiment. All of the participants were native Turkish speakers. None of the participants reported any hearing-related or vocal problems. Also, none of the participants had any prior experience of listening experiments.

One of the couples from the originally employed group decided not to take part in the listening experiment and was removed from the set of participants. However, the speech signals obtained from those subjects were retained for the listening experiment. Therefore, a total of 22 participants (11 male and 11 female with an age range of 21–30 years) took part in the listening experiment.

### 2.2 Stimuli

Stimuli consisted of binaural emulations of different sound source distances obtained by convolving dry speech signals from the participants with binaural room impulse responses measured in a reverberant skyway.

#### 2.2.1 Speech signals

Speech signals used to produce the binaural stimuli were recorded in the METU Spatial Audio Research Group (SPARG) Lab which has a floor area of 11 m<sup>2</sup> and a low reverberation time of  $T_{30} = 80$  ms. The background noise within the lab is below 40 dB SPL. The structure is specially designed with oblique walls to avoid strong standing waves. The walls are covered with 10 cm thick mineral wool screened with acoustically transparent textile to reduce reflections. Therefore, the recordings were sufficiently dry and were treated as being practically anechoic as far as the present experimental conditions are concerned.

Five different sentences in Turkish language were constructed using words that had been scaled as being emotionally neutral [30]. These sentences and their English translations are:

- S1: *Dünya gezegendir.*  
(EN: Earth is a planet.)
- S2: *Benden kazak aldı.*  
(EN: S/he took a sweater from me.)
- S3: *Ormanda ağaç var.*  
(EN: There are trees in the forest.)
- S4: *O, radyoyu açtı.*  
(EN: S/he turned the radio on.)
- S5: *Kaşık çekmecede.*  
(EN: The spoon is in the drawer.)

Note that all of these sentences have an equal number of six syllables and are thus balanced. Note also that since

Turkish is a gender-neutral language, these sentences do not contain any gender-specific information.

Recordings were made using a cardioid vocal microphone (BM-800) positioned 1 m away from the speaker. The height of the microphone and its acoustic axis were adjusted to match the speaker's mouth before each recording. The speakers were instructed to maintain their regular speaking level avoiding any additional vocal effort and emotional prosody during the recordings. The sampling rate was 44.1 kHz and amplitude resolution was 32-bits. There were two reasons for recording speech at a distance from the speaker and not using a miniature microphone positioned close to the speaker's mouth: (1) in order to reduce the *proximity effect* that accentuates low frequencies due to recording in the near-field of the sound source [31], and (2) in order to provide a visual anchor according to which the speakers can adjust their voices.

In order to assess the effects of interpersonal familiarity, speech signals to be used in the experiment were selected differently for each participant. We defined three familiarity categories for this purpose: (1) *Familiar* (V0) represents the sentence as spoken by the participant's partner, (2) *Similar–Unfamiliar* (V1) corresponds to the same sentence as spoken by an unfamiliar speaker with the highest long-term spectral similarity, and (3) *Dissimilar–Unfamiliar* (V2) corresponds to the same sentence as spoken by an unfamiliar speaker with the lowest long-term spectral similarity.

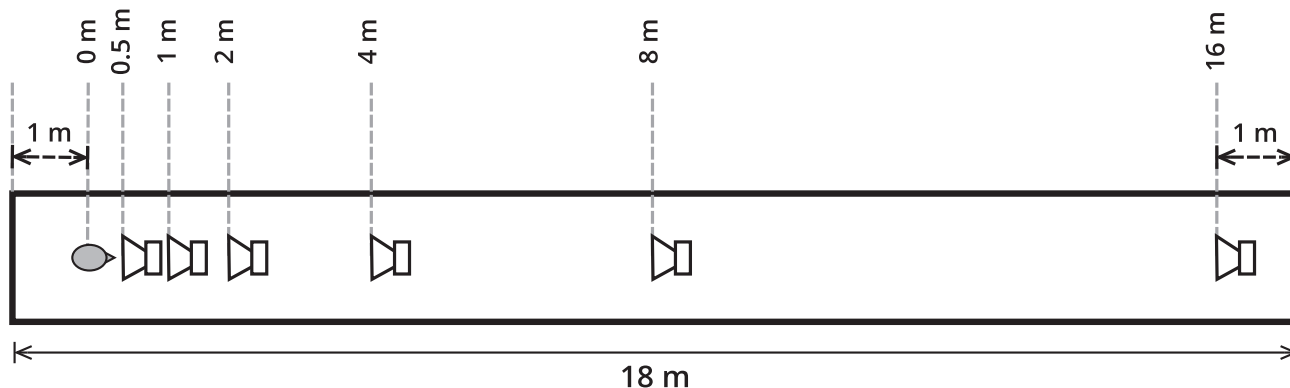
The selection of speakers for the V0 case were obviously uniform across all participants (i.e. familiar speaker for each participant was their partner), while the selection of V1 and V2 were not. The female and male speakers most frequently selected for the V1 case were fm06 (10/60), fm02 (7/60) and fm10 (6/60), and ml10 (12/60), ml06 (7/60) and ml07 (7/60), respectively. The female and male speakers most frequently selected for the V2 case were fm08 (22/60), fm09 (13/60) and fm11 (11/60), and ml08 (13/60), ml03 (12/60) and ml07 (7/60), respectively. While there is some imbalance in this selection, a single speaker does not overwhelmingly dominate the results.

The long-term spectral similarity was calculated using the Euclidean distances between normalized vectors representing the band energies of speech signals processed using a 64 channel Gammatone filterbank with center frequencies matching the center frequencies of equivalent rectangular bandwidth (ERB) filters [32].

Since the participants would have been familiar with the presented stimuli after the experiment, a repetition of the experiment with the same set of participants would not have been possible.

#### 2.2.2 Binaural stimuli

An observational study of auditory distance perception is in general not feasible due to the difficulties involved in controlling the experimental conditions. A practical alternative involves presenting the subjects with binaural recordings that emulate natural spatial hearing scenarios with the stimuli presented over a pair of headphones.



**Figure 1.** The measurement positions shown on the top view of the skyway. The gray ellipse shows the position of the binaural microphone and the positions of sound sources are denoted by loudspeaker symbols.

In order to achieve flexibility in the creation of different stimuli corresponding to different source distances, binaural room impulse responses (BRIRs) were measured in the skyway connecting the two wings of the METU Graduate School of Informatics. The dimensions of the skyway are  $18 \times 1.8 \times 2.5$  m. The sides of the skyway are glass, the floor is marble and the ceiling is made of curved corrugated metal. The reverberation time was measured as  $T_{30} = 1.15$  s.

A binaural microphone (Neumann KU-100) was placed at a height of 1.5 m and on the midline 1 m away from one end of the skyway. BRIRs in the front direction of the binaural microphone were measured using a small, two-way active loudspeaker (Genelec 6010A) positioned 0.5, 1, 2, 4, 8, and 16 m away from the binaural microphone at 1.5 m height as shown in Figure 1. The decision to select these particular distances is motivated by an earlier study of auditory distance perception using a similar logarithmic spacing [23]. The BRIRs were measured using the logarithmic sine sweep method [33].

A total of 90 stimuli per participant (5 sentences, 6 distances, 3 familiarity categories) were prepared by convolving the previously recorded dry speech signals with the measured BRIRs. This way, each participant received a different set of stimuli.

## 2.3 Procedure

### 2.3.1 Level equalization

The recording levels in the nearly-anechoic conditions under which the dry speech signals were captured were similar but not identical across participants. Similarly, the energy of the late reverberation was assumed to be largely the same for all recording distances. The presentation levels of the stimuli were therefore equalized to eliminate the possible confounding effect of level. Considering earlier studies showing that D/R ratio is a distance cue that can be as effectively used as level especially at high reverberation levels [27], our assumption was that reliable responses can be elicited with level-equalized stimuli.

Level equalization resulted in stimuli with similar total energy whilst substantially preserving the change in the

level of the direct sound according to the inverse-square law, consequentially preserving the relative direct path level cues and the relative D/R ratios with respect to the source distance. The motivation behind this choice was to prevent the participants from adapting to the levels of the presented stimuli. In other words, level equalization served to eliminate the dominant effect of level cues, which may have confounded the responses since level is both a physical (i.e. due to sound propagation) and a cognitive (i.e. due to familiarity with the speaker’s nominal speech level) cue. Therefore, the participants were forced, in effect, to use D/R ratio as the only available dominant distance cue, making the differences observed in the elicited responses attributable either to D/R ratio differences or to the differences in interpersonal familiarity.<sup>1</sup> The physical effects of level equalization on the relative levels of direct sound and reverberation are discussed in the Appendix.

Level equalization was carried out by normalizing binaural stimuli for their total broadband energy with respect to the highest total energy across the employed left and right ear signals. The presentation level was then selected in the following way. A 1 kHz sinusoidal signal was played back from the headphones employed in the listening test (Superlux HD-330) and the reproduction level was measured using a calibrated miniature microphone positioned at the entrance of the right ear canal of the binaural microphone using a sound level meter (Faber Acoustical Sound Meter). The level of the headphone amplifier (Focusrite Scarlett 18i8) was adjusted to and fixed at 80 dB SPL. The processed speech signals were then played back via the same pair of headphones and the sound level was measured. The average peak and equivalent sound pressure levels,  $L_p$  and  $L_{eq}$ , were 77.3 dB and 66.1 dB for female speech samples and 77.5 dB and 65.9 dB for male speech samples, respectively. These values are generally in agreement with the average level of running speech [34]. While the presentation level varied slightly for each stimulus, the variance of  $L_{eq}$  was less than about 2 dB for most cases.

<sup>1</sup> As will be discussed later, it was observed that semantic content also modulated distance judgements, an effect that was not foreseen.

### 2.3.2 Experimental task

The 22 participants whose voices were recorded earlier also took part in the listening experiment. Each participant attended the experiment alone. Level equalized stimuli were presented over headphones in the METU SPARG Lab. The participant's task was to listen to the presented stimulus and to adjust a vertical slider to indicate the perceived distance. The participants were allowed to listen to each stimulus only once. The end points of the slider represented 0 and 20 m distance. The reason for selecting an upper limit of the response distance was based on our assumption that source distances would be underestimated on average and especially so for sources at larger distances. OpenSesame software [35] was used to set up, control and run the experiment. In order to alleviate the effects of visual feedback on distance perception, the room was darkened except for the light coming from the computer screen. The stimuli presented to the participants were randomized for distance, sentence and familiarity and each stimulus was presented only once. The participants were not informed about how the stimuli were obtained. Overall, each participant provided 90 responses. The experiment took approximately 10 min on average per participant to complete.

A training session was run for each participant prior to the actual experiment in order to acquaint them with the user interface as well as the experimental task. Binaural stimuli produced with speech signals different from the ones used in the experiment were used during the training session. In order to demonstrate the range of stimuli that the participants will hear, three examples from the nearest (i.e. 0.5 m) and three examples from the furthest (i.e. 16 m) distance were presented. No feedback was provided about the actual distances of the stimuli. The stimuli were assumed to be well-externalized since none of the participants reported inside-the-head localization.

After the experiment, one of the participants stated that he did not fully understand the experimental task. The data collected from that participant were discarded.

## 3 Results

Following methods used in previous auditory distance perception studies [1, 2, 6, 7, 36]; a compressive power function proposed by Zahorik [7] was fit to collected data. The representation defines the relation between the actual distance,  $D$  and the perceived distance  $\bar{D}$  as a compressive power function in the following form:

$$\bar{D} = \beta D^\alpha, \quad (1)$$

where  $\beta$  is a constant and  $\alpha$  the power law exponent. The compressive model parameter values present a high-level summary of the collected responses. The parameters,  $\alpha$  and  $\beta$ , of the compressive model represent how compressive the auditory distance perception is, and how much the distance of a source positioned at 1 m is over- or underestimated, respectively. The ideal case where the simulated and the perceived distances are equal corresponds to

$\alpha = \beta = 1$ . A higher value of  $\alpha < 1$  corresponds to a lower compression resulting in a higher slope in the log perceived distance vs. log actual distance line. A higher value of  $\beta > 1$  corresponds to a higher overestimation of the distance of a source positioned at 1 m and vice versa.

After eliminating the outliers (defined as a data point outside 1.5 times the interquartile range after logarithmic transform of the data), the geometric means of the collected responses were calculated and least-squares fits of power functions as described above were obtained. Figure 2 shows the aggregate responses for each participant as well as the power function fits to geometric means of responses for each simulated distance.

It was observed that some of the participants had very low  $\alpha$  values resulting in a flat curve indicating either that (i) the participant responded completely randomly or (ii) they substantially and consistently provided similar responses regardless of the stimulus. Either situation is indicative of a problematic response pattern. Therefore, responses from participants with an  $\alpha$  that is one standard deviation below the mean were discarded from the analyses. This corresponds to 4 participants (P16, P17, P18, P19) over the entire set of 21 participants. The mean values of the fitted parameters of the remaining 17 participants were  $\alpha = 0.31 \pm 0.06$  and  $\beta = 3.56 \pm 1.12$ . The average coefficient of determination was  $R^2 = 0.82$  indicating a good overall fit.

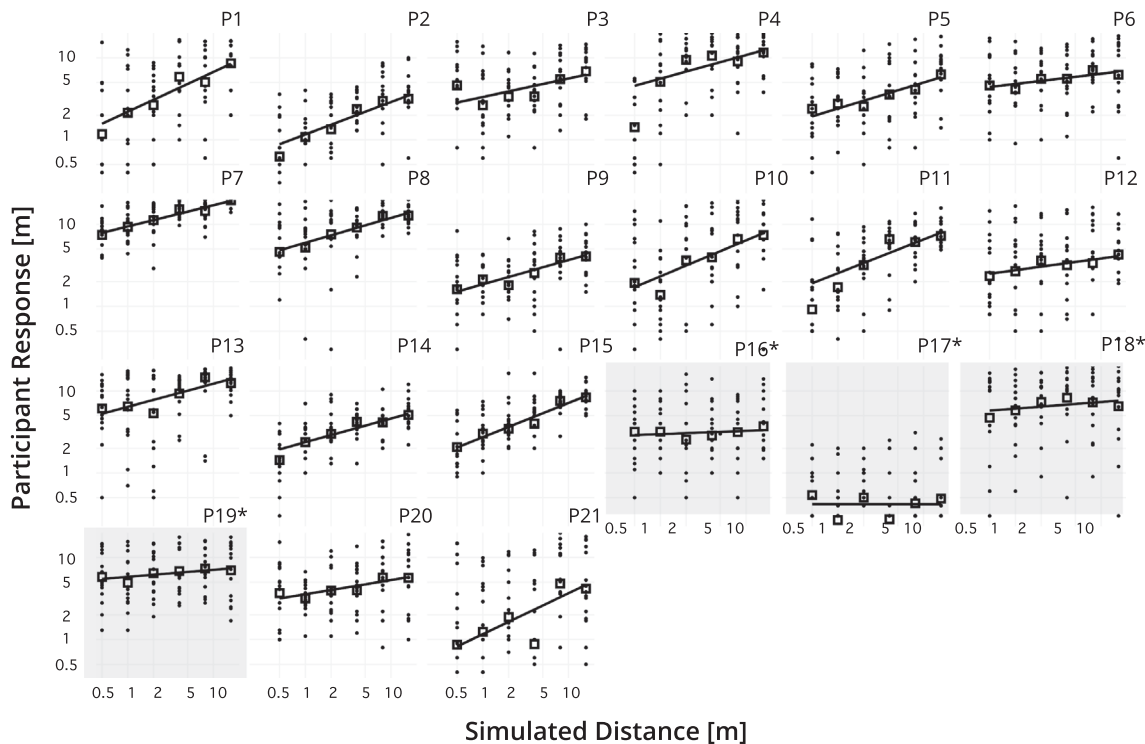
Figure 3 shows the aggregated responses for different familiarity conditions. Points indicate individual responses. Square symbols indicate geometric means. The dashed diagonal line represents the ideal, veridical response where  $D = \bar{D}$ . Data points above this line correspond to overestimated responses and vice versa. The compressive functions fit to geometric means are also shown. The fitted values of  $\alpha$  and  $\beta$  are indicated on the figures for different familiarity levels. The values of  $\alpha$  and  $\beta$  differ slightly across different levels of familiarity where  $\alpha$  is larger and  $\beta$  is smaller than both unfamiliar conditions for the familiar condition.

As a second step, an analysis of variance (ANOVA) was used to examine the effects of the tested factors on distance perception. The factors are *Familiarity*, *Sentence* and *Distance*. The dependent variable is the log-distance error defined as:

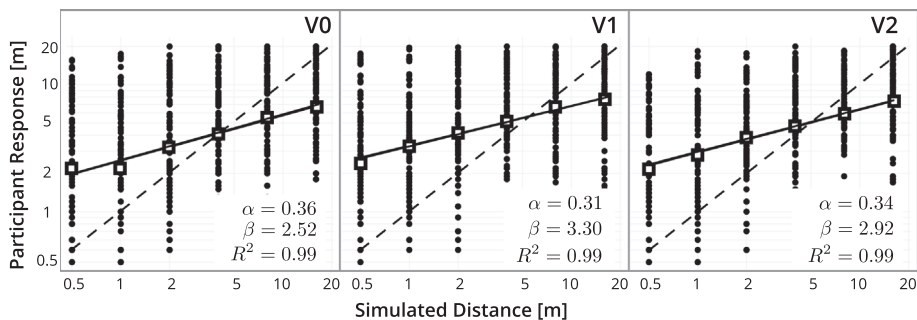
$$E = \log D - \log \bar{D}. \quad (2)$$

When  $E < 0$ , the source distance is overestimated in comparison with the actual distance (i.e.  $\bar{D} > D$ ) and vice versa. Also,  $E \approx 0$  indicates veridical localization of distance since (i.e.  $\bar{D} \approx D$ ). In other words, a response with a small absolute log-distance error is considered more veridical. Notice that logarithmic transformation of the perceived and actual distances were used to normalize the variance of the error across different simulated distances. The log-distance errors obtained this way were confirmed via individual Kolmogorov–Smirnov tests [37] to be normally distributed, thereby allowing the use of an ANOVA model for further analysis.

The statistical model used for analyzing the elicited responses was a  $3 \times 5 \times 6$  ANOVA where the main factors



**Figure 2.** Perceived distances as functions of simulated distances for different participants. Each point indicates a single participant response. Squares represent geometric means. The solid curves show the power functions fitted to geometric means. Participants whose responses were eliminated from further analysis are shown with gray boxes.



**Figure 3.** Perceived distances as functions of simulated distances for different familiarity levels (V0, V1, and V2). Parameters of the fitted power functions ( $\alpha$  and  $\beta$ ) as well as coefficients of determination ( $R^2$ ) are also shown. The dashed diagonal line represents veridical distance localization (i.e.  $D = \bar{D}$ ).

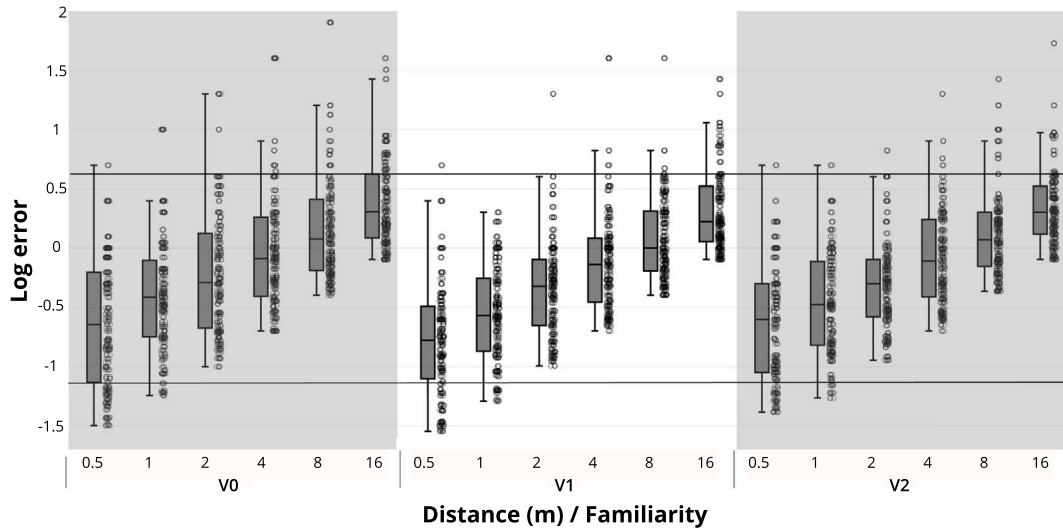
were *Familiarity*, *Sentence*, and *Distance*. Since an initial analysis revealed that none of the interaction terms was significant except for *Familiarity*  $\times$  *Sentence*, these terms were removed in order to increase the statistical power of the employed model.

ANOVA indicates that differences between the mean log-distance errors are statistically significant for *Distance* [ $F(5, 1510) = 187.924, p < 0.001, \omega^2 = 0.388$ ] with a large effect size, *Familiarity* [ $F(2, 1510) = 8.867, p < 0.001, \omega^2 = 0.006$ ] and *Sentence* [ $F(4, 1510) = 15.540, p < 0.001, \omega^2 = 0.022$ ] with small effect sizes. The two-way interaction term, *Familiarity*  $\times$  *Sentence* was also found to be statistically significant [ $F(8, 1510) = 2.4,$

$p = 0.014, \omega^2 = 0.004$ ]. As expected, most of the variability in the data is explained by the factor *Distance*.

Post-hoc comparisons were carried out by bootstrapping with 1000 repetitions and using the Bonferroni correction. Post-hoc comparisons for different sentences show that the differences of log-distance error between S5 and the other sentences are statistically significant ( $p < 0.001$ ). In all cases average log-distance error difference is negative, indicating that S5 is localized at a closer distance on average than other sentences.

Post-hoc comparisons of different familiarity levels revealed that the differences between the log-distance error means of the familiarity levels, V0 and V1 ( $M = 0.111,$



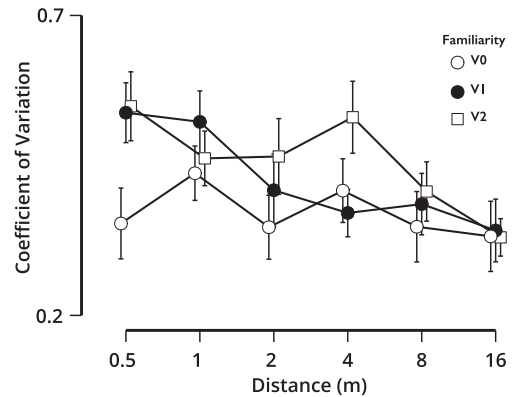
**Figure 4.** Log-distance errors for different distances and familiarity levels.

$SE = 0.026$ ) are statistically significant ( $p < 0.001$ ) whereas the difference between V0 and V2 ( $M = 0.053$ ,  $SE = 0.028$ ) as well as V1 and V2 ( $M = -0.057$ ,  $SE = 0.025$ ) are not statistically significant. The effect size for the V0–V1 contrast was  $d = 0.197$ , which is small according to Cohen [38]. Figure 4 shows the log-distance errors for different familiarity levels and distances.

The differences between the mean log-distance errors are statistically significant across all pairs of distances with  $p < 0.001$ . In all cases, the effect sizes ranged from medium to large. The absolute log-distance error linearly increased with increasing actual log-distance, indicating a higher level of underestimation at larger distances.

An interesting finding from the post-hoc comparison of different familiarity-sentence pairs is that S5 when spoken by the familiar speaker, (V0, S5), is localized significantly closer ( $p < 0.001$ ) than other sentences (S1–S4) not only for the familiar speaker for the combinations (V0, S2), (V0, S3), (V0, S4) but also for the similar-unfamiliar speaker for combinations (V1, S1), (V1, S2), (V1, S3), (V1, S4) and dissimilar-unfamiliar (V2) speaker for the conditions (V2, S1), (V2, S2), and (V2, S3). The only exceptions are the differences between (V0, S5) and the combinations, (V0, S1) which is marginally insignificant ( $p = 0.063$ ), and (V2, S4) which is not significant. No other systematic differences were observed between other combinations.

Another question related to distance perception accuracy is about the effect of familiarity on the variability in the elicited responses. We used *coefficient of variability* (CV), defined as the ratio of the standard deviation to the mean, in order to assess whether the responses given to different levels of familiarity differed in their dispersions (Fig. 5). CV was calculated separately for each participant and for all sentences over responses grouped according to distance and familiarity level. ANOVA was used to investigate the possible differences between values of CV calculated for different experimental conditions. ANOVA



**Figure 5.** Coefficient of variation (CV) for different sentences and different familiarity levels. Mean values and standard error are shown.

revealed that only *Distance* was a significant factor [ $F(5, 286) = 3.604$ ,  $p = 0.004$ ,  $\omega^2 = 0.042$ ]. Bootstrapped post-hoc comparisons with 1000 repetitions and Bonferroni correction revealed that CVs were different only between  $D = 0.5$  m and  $D = 16$  m ( $p = 0.002$ ) where CV was smaller for responses given to stimuli at 16 m which may be due to the upper limit of the employed scale.

## 4 Discussion

The main finding of the presented study is that interpersonal familiarity is a factor that can modulate the auditory perception of distance of reverberant speech. A more detailed discussion of this and other findings follows.

### 4.1 Main findings

The compressive model parameter values obtained using the aggregate responses of each participant indicate that for nearby sound sources, especially within a distance

less than about 3.5 m on average, participants tended to overestimate the source distance, while they underestimated distances of sources that are further away. This is slightly different than earlier results in the literature where subjects start to overestimate distance when the source distance is less than about 1 m. This difference is very likely to occur due to the employed level equalization which results in approximately 3 dB difference between the actual and the normalized direct path components for source distances greater than 1 m (see Fig. A.1 in Appendix). In other words, level-equalized stimuli have a lower D/R ratio than natural stimuli for distances larger than the critical distance of the room. Therefore, source distances are overestimated more on average than in studies that employ non-equalized, naturalistic stimuli [36, 39] since distance perception substantially relies on the D/R ratio as the predominant distance cue in the absence of absolute level cues.

The compressive model parameter values we obtained can be compared with the results of recent studies employing a similar binaural presentation approach. In one of these studies, Anderson and Zahorik [36] used windowed Gaussian noise convolved with BRIRs measured in a concert hall with a reverberation time of  $T_{60} = 1.9$  s and reported an average  $\alpha$  of  $0.61 \pm 0.30$  and an average  $\beta$  of  $2.22 \pm 1.99$ . In another study using anechoic speech signals convolved with synthetic BRIRs, increasing the reverberation time was found to increase distance estimation errors and average parameter values of  $\alpha = 0.53$  and  $\beta = 1.89$  were reported for a reverberation time of  $T_{60} = 0.86$  s [39]. In comparison with both of these studies, the parameter values observed in the reported experiment indicate more compression (i.e. lower  $\alpha$  values) and a higher distance offset (i.e. higher  $\beta$  values). These differences might be due to the employed stimuli, the elongated shape and the resulting acoustic features (e.g. lateral reflections) of the skyway, level equalization, or the lack of vocal effort as a covariate factor in the experimental design.

Familiar speakers (V0) were on average judged to be closer than unfamiliar speakers that have the most similar long-term spectra (V1). This difference was observed both as differences between the compressive model parameter values of these two conditions and via post-hoc comparisons applied on the log-distance errors.

Based on the compressive function fits to aggregated responses, the value of the parameter  $\alpha$  was highest for V0 ( $\alpha = 0.36$ ), followed by V2 ( $\alpha = 0.34$ ) and V1 ( $\alpha = 0.31$ ), respectively. In a similar way  $\beta$  was lowest for V0 ( $\beta = 2.52$ ) followed by V2 ( $\beta = 2.92$ ) and V1 ( $\beta = 3.30$ ), respectively. In other words, the familiar speaker is on average judged to be closer and more accurately localized in distance than an unfamiliar speaker, irrespective of its spectral similarity. These results indicate the possible role of long-term exposure to a familiar speaker's voice in increasing subjective distance localization accuracy. The higher level of the value of the parameter,  $\beta$ , for unfamiliar speakers can be linked to differences in proxemic behavior towards familiar and unfamiliar people, where a smaller interpersonal distance is typically preferred for the former and a larger interpersonal distance for the latter [40, 41].

The statistical analysis of the log-distance errors revealed that the difference between the log-distance error means is statistically significant for V0 and V1.<sup>2</sup> In contrast, no statistically significant difference exists between either familiar (V0) and unfamiliar speakers with dissimilar long-term spectra (V2) or the latter with similar (V1) long-term spectra. This is an unexpected result considering the assumption that V1 would constitute an intermediate step between V0 and V2 in some familiarity continuum. This mismatch between our expectation and the observed results might possibly be due to the spectrally most similar speaker being judged as more unfamiliar than the dissimilar–unfamiliar speaker. Such a difference can be attributed to the *perceptual magnet effect* which occurs as the perception of a stimulus is modulated by the categorical perception of that stimulus as familiar or unfamiliar. A similar–unfamiliar stimulus closer to a familiar/unfamiliar decision boundary can be perceived as being even more unfamiliar than a dissimilar–unfamiliar stimulus [42, 43]. It is likely that familiarity with other features of speech signals such as vocal effort, rhythm, intonation and other prosodic features can also have analogous effects to spectral similarity in modulating distance judgements. A future study might be envisaged, that uses speech signals generated via morphing between the familiar and dissimilar–unfamiliar speaker voices to provide a clearer understanding of the effect not only of spectral similarity but also of other prosodic features on auditory distance perception, possibly also taking into account the *perceptual magnet effect*.

Another finding concerns the effect of semantic content as observed from the statistically significant differences between mean log-distance errors of S5 and other sentences. We speculate that the contextual cues in S5 calling for joint attention (i.e. *to a spoon*) in close vicinity (i.e. *in the drawer*) in simple present tense might have resulted in an underestimation of distance judgements for that sentence on average. This finding points at the possible modulating effect of semantic cues on the auditory distance perception of reverberant speech.

Finally, the distance of the stimulus S5 when uttered by the familiar speaker is perceived to be closer than combinations of other sentences with other speakers, familiar or unfamiliar. This finding is a possible evidence of the combined effect of familiarity and semantic content in modulating distance perception.

## 4.2 Limitations

The employed experimental design allows tightly controlling the experimental variables. However, it has also likely resulted in level, D/R ratio, vocal effort combinations that do not necessarily correspond to real-life listening scenarios. For example, familiarity with the expected vocal effort combined with visual cues has a strong effect on distance perception in a conversational context [36]. However,

<sup>2</sup> It should be noted that despite the observed statistical significance, the effect size for V0–V1 contrast is small. In other words, interpersonal familiarity modulates the auditory distance perception but is not a strong distance cue, per se.



this study did not include vocal effort as a factor or covariate.

Level equalization resulted in the modification of absolute level cues, which could have been used by the participants as reference points. The lack of overall level as a cue is one of the possible reasons for the increased variability in the elicited responses in comparison with earlier studies (e.g. [2]). Similarly, the increased overestimation of source distance beyond the room's critical distance can be attributed to the non-linear change in the direct path level (see Fig. A.1 in the Appendix).

Another possible limitation concerns the finite upper limit of the employed scale. Some of the participants used the full scale and judged some of the stimuli to be further away than 16 m, which is where the furthest source emulated in the experiment was positioned. This upper limit might have biased the results, especially for larger stimulus distances. Similarly, a source discrimination or identification task instead of the employed absolute distance task could have also decreased the response variability by rendering the experiment easier for the participants.

Another possible limitation is that the participants themselves did not make similarity judgements for the presented stimuli in order to ensure that they were not familiarized with the employed stimuli prior to the experiment. It is possible that more appropriate ways than Euclidean distance based approach can be found to quantify spectral similarity. For example, an *A*-weighted distance metric to account for the frequency-dependent sensitivity of the human auditory system could be more appropriate.

## 5 Conclusions

We presented the results of a binaural listening experiment where the participants judged the distances of familiar and unfamiliar speakers using level-equalized stimuli. The experiment involved the headphone based presentation of five balanced and nominally neutral sentences spoken by different-gender couples, convolved with BRIRs recorded in an elongated skyway. Each participant listened to speech signals from three different participants including their long-term partner and two unfamiliar people of opposite gender with the lowest and highest long-term spectral similarity to the participant's partner, respectively. The participants then judged the apparent distances of the stimuli that they listened to, using a scale with 0 and 20 m as its endpoints.

The study showed that interpersonal familiarity modulates auditory distance perception. Distance of a binaurally presented reverberant speech signal from a familiar speaker is on average more accurately localized than that from an unfamiliar speaker. However, the difference is only statistically significant between familiar and similar-unfamiliar conditions. The study revealed another interesting finding that is counterintuitive at first sight: the distance of a spectrally similar and unfamiliar speech signal is localized even further than the spectrally dissimilar and unfamiliar signal. We speculate that this finding is related to the *perceptual magnet effect*.

To the best of the authors' knowledge this is the first study in literature investigating the effects of interpersonal familiarity on auditory distance perception. Long-term familiarity is very likely to modulate distance perception also under natural listening conditions, and possibly more prominently than in the experiment we reported, since additional cues such as absolute level, vocal effort and prosody would also be available. From an applied perception point of view, the findings can have practical implications in the domains of human–robot or human–agent interaction as well as the design of shared virtual and mixed reality environments.

## Conflict of interest

Authors declared no conflict of interests.

## Acknowledgments

We thank Professor Cem Bozşahin for useful discussions on an earlier version of the presented work.

## References

1. J. Blauert: Spatial Hearing: The Psychophysics of Human Sound Localization. MIT Press, Cambridge, MA, 1997.
2. P. Zahorik: Auditory distance perception in humans: A summary of past and present research. Acta Acustica United with Acustica 91, 3 (2005) 409–420.
3. H. Kuttruff: Room Acoustics. CRC Press, 2016.
4. A.J. Kolarik, B.C.J. Moore, P. Zahorik, S. Cirstea, S. Pardhan: Auditory distance perception in humans: A review of cues, development, neuronal bases, and effects of sensory loss. Attention, Perception & Psychophysics 78, 2 (2015) 373–395.
5. A.W. Bronkhorst, T. Houtgast: Auditory distance perception in rooms. Nature 397 (1999) 517–520.
6. P. Zahorik: Assessing auditory distance perception using virtual acoustics. Journal of the Acoustical Society of America 111, 4 (2002a) 1832–1846.
7. P. Zahorik: Direct-to-reverberant energy ratio sensitivity. Journal of the Acoustical Society of America 112, 5 (2002b) 2110–2117.
8. E. Larsen, N. Iyer, C.R. Lansing, A.S. Feng: On the minimum audible difference in direct-to-reverberant energy ratio. Journal of the Acoustical Society of America 124, 1 (2008) 450–461.
9. A.D. Little, D.H. Mershon, P.H. Cox: Spectral content as a cue to perceived auditory distance. Perception 21, 3 (1992) 405–416.
10. L.B. Evans, H.E. Bass, L.C. Sutherland: Atmospheric absorption of sound: Theoretical predictions. Journal of the Acoustical Society of America 51, 5B (1972) 1565–1575.
11. M.B. Gardner: Distance estimation of 0° or apparent 0°-oriented speech signals in anechoic space. Journal of the Acoustical Society of America 45, 1 (1969) 47–53.
12. E.C. Healey, R. Jones, R. Berky: Effects of perceived listeners on speakers' vocal intensity. Journal of Voice 11, 1 (1997) 67–73.
13. P. Zahorik, J.W. Kelly: Accurate vocal compensation for sound intensity loss with increasing distance in natural

- environments. *Journal of the Acoustical Society of America* 122, 5 (2007) EL143–EL150.
14. H. Traunmüller, A. Eriksson: Acoustic effects of variation in vocal effort by men, women, and children. *Journal of the Acoustical Society of America* 107, 6 (2000) 3438–3451.
  15. J.W. Philbeck, D.H. Mershon: Knowledge about typical source output influences perceived auditory distance. *Journal of the Acoustical Society of America* 111, 5 (2002) 1980–1983.
  16. A. Eriksson, H. Traunmüller: Perception of vocal effort and distance from the speaker on the basis of vowel utterances. *Perception & Psychophysics* 64, 1 (2002) 131–139.
  17. D. Pelegrín-García, B. Smits, J. Brunskog, C.-H. Jeong: Vocal effort with changing talker-to-listener distance in different acoustic environments. *Journal of the Acoustical Society of America* 129, 4 (2011) 1981–1990.
  18. J.M. Loomis, R.L. Klatzky, J.W. Philbeck, R.G. Golledge: Assessing auditory distance perception using perceptually directed action. *Perception & Psychophysics* 60, 6 (1998) 966–980.
  19. E.R. Calcagno, E.L. Abregu, M.C. Eguía, R. Vergara: The role of vision in auditory distance perception. *Perception* 41, 2 (2012) 175–192.
  20. K.T. Gagnon, M.N. Geuss, J.K. Stefanucci: Fear influences perceived reaching to targets in audition, but not vision. *Evolution and Human Behavior* 34, 1 (2013) 49–54.
  21. M.G. Wisniewski, E. Mercado, K. Gramann, S. Makeig: Familiarity with speech affects cortical processing of auditory distance cues and increases acuity. *PLoS One* 7, 7 (2012) e41025.
  22. P.D. Coleman: Failure to localize the source distance of an unfamiliar sound. *Journal of the Acoustical Society of America* 34, 3 (1962) 345–346.
  23. D.S. Brungart, K.R. Scott: The effects of production and presentation level on the auditory distance perception of speech. *Journal of the Acoustical Society of America* 110, 1 (2001) 425–440.
  24. J.J. Zwillocki: *Sensory Neuroscience: Four Laws of Psychophysics*. Springer Science + Business Media, Syracuse, NY, 2009.
  25. D.H. Mershon, J.N. Bowers: Absolute and relative cues for the auditory perception of egocentric distance. *Perception* 8, 3 (1979) 311–322.
  26. M.A. Akeroyd, S. Gatehouse, J. Blaschke: The detection of differences in the cues to distance by elderly hearing-impaired listeners. *Journal of the Acoustical Society of America* 121, 2 (2007) 1077–1089.
  27. A. Kolarik, S. Cirstea, S. Pardhan: Discrimination of virtual auditory distance using level and direct-to-reverberant ratio cues. *Journal of the Acoustical Society of America* 134, 5 (2013) 3395–3398.
  28. 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. *Experimental Brain Research* 224, 4 (2013) 623–633.
  29. A. Bidart, M. Lavandier: Room-induced cues for the perception of virtual auditory distance with stimuli equalized in level. *Acta Acustica United with Acustica* 102, 1 (2016) 159–169.
  30. D. Gökçay, M.A. Smith: TÜDADEN: Türkçe’de duygusal ve anlamsal değerlendirmeli norm veri tabanı (TUDADEN: Database of emotionally and semantically validated norms in Turkish language), in: *Bilgisayar ve Beyin H. Bingöl, Editors Istanbul, Pan Yayınck. 2012.*
  31. M. Kleiner: *Electroacoustics*. CRC Press, 2013.
  32. M. Cooke: *Modelling Auditory Processing and Organisation*. Cambridge University Press, Cambridge, UK, 1993.
  33. A. Farina: Simultaneous measurement of impulse response and distortion with a swept-sine technique, in *Proc. 108th Audio Eng. Soc. Conv., number preprint #5093, Paris, France. 2000.*
  34. P. Corthals: Sound pressure level of running speech: percentile level statistics and equivalent continuous sound level. *Folia Phoniatrica et Logopaedica* 56, 3 (2004) 170–181.
  35. S. Mathôt, D. Schreij, J. Theeuwes: OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods* 44, 2 (2012) 314–324.
  36. P.W. Anderson, P. Zahorik: Auditory/visual distance estimation: Accuracy and variability. *Frontiers in Psychology* 5, 1097 (2014) 1–11.
  37. F.J. Massey Jr: The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American Statistical Association* 46, 253 (1951) 68–78.
  38. J. Cohen: *Statistical Power Analysis for the Behavioural Sciences*. Erlbaum, Hillsdale, NJ, 1988.
  39. M. Paquier, N. Côté, F. Devillers, V. Koehl: Interaction between auditory and visual perceptions on distance estimations in a virtual environment. *Applied Acoustics* 105 (2016) 186–199.
  40. N.L. Ashton, M.E. Shaw, A.P. Worsham: Affective reactions to interpersonal distances by friends and strangers. *Bulletin of the Psychonomic Society* 15, 5 (1980) 306–308.
  41. J.W. Burgess: Interpersonal spacing behavior between surrounding nearest neighbors reflects both familiarity and environmental density. *Ethology and Sociobiology* 4, 1 (1983) 11–17.
  42. P.K. Kuhl: Human adults and human infants show a “perceptual magnet effect” for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics* 50, 2 (1991) 93–107.
  43. R.K. Moore: A Bayesian explanation of the “uncanny valley” effect and related psychological phenomena. *Scientific Reports* 2, 1 (2012) 1–5.

## Appendix

### Level equalization

We will first show that the total energy of a signal convolved with a room impulse response is proportional to the sum of energies of the direct path and reverberation components. We will then show that the effect of energy normalization over the level of direct path component does not substantially change the attenuation of the direct path with respect to distance.

Let us assume that a sound signal  $x(n)$  is convolved with a (binaural) room impulse response  $h(n)$  to generate an output  $y(n) = x(n) * h(n)$  where the room impulse response comprises a direct path component and a reverberation component such that:

$$h(n) = h_d(n) + h_r(n), \quad (\text{A.1})$$

where  $h_d(n)$  is the direct path component and  $h_r(n)$  is the reverberation component. The energy of the signal  $y(n)$  can be expressed as:

$$\epsilon_y = \sum_n |y(n)|^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |Y(\omega)|^2 d\omega, \quad (\text{A.2})$$

using Parseval's relation, where  $\omega$  is the angular frequency. Let us define a simplified impulse response as a combination of a direct path component and a reverberation component such that:

$$h(n) = \alpha_d \delta(n) + h_r(n), \quad (\text{A.3})$$

where  $\alpha_d \in \mathbb{R}$  is the amplitude of the direct path component, and  $\delta(n)$  is the unit sample signal. Let us assume that  $h_r(n)$  can be modelled as exponentially decaying white Gaussian noise such that:

$$h_r(n) = e^{-kn} \bar{R}(n), \quad (\text{A.4})$$

where  $\bar{R}(n)$  comes from a normal distribution such that  $\bar{R}(n) \sim \mathcal{N}(0, \sigma_r^2)$  where  $\sigma_r^2$  is the variance.

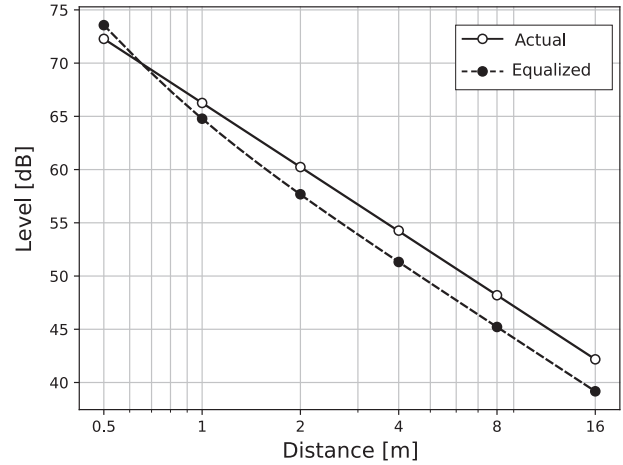
We can express the energy density spectrum of the output in frequency domain as:

$$\begin{aligned} |Y(\omega)|^2 &= |X(\omega)|^2 |H_d(\omega) + H_r(\omega)|^2 \\ &= |X(\omega)|^2 (\alpha_d^2 + 2\alpha_d \Re[H_r^*(\omega)] + |H_r(\omega)|^2), \end{aligned} \quad (\text{A.5})$$

using  $H_d(\omega) = \alpha_d$ . The output energy is:

$$\begin{aligned} E\{\epsilon_y\} &= \frac{1}{2\pi} \int_{-\pi}^{\pi} E\{|X(\omega)|^2 (\alpha_d^2 + |H_r(\omega)|^2)\} d\omega \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} E\{|X(\omega)|^2\} (\alpha_d^2 + E\{|H_r(\omega)|^2\}) d\omega, \end{aligned} \quad (\text{A.6})$$

where  $E\{\cdot\}$  is the expectation operator and we used  $E\{\Re[H_r(\omega)]\} = 0$  and statistical independence of  $X(\omega)$  and  $H_r(\omega)$ . This way, it becomes possible to define the energy contributions of the direct path and the reverberant part separately, such that:



**Figure A.1.** Levels of direct path components for actual and level-equalized cases.

$$\epsilon_{y,d} = \frac{1}{2\pi} \int_{-\pi}^{\pi} E\{|X(\omega)|^2\} \alpha_d^2 d\omega, \quad (\text{A.7})$$

$$\epsilon_{y,r} = \frac{1}{2\pi} \int_{-\pi}^{\pi} E\{|X(\omega)|^2\} E\{|H_r(\omega)|^2\} d\omega. \quad (\text{A.8})$$

Given the definition in (A.4), it is possible to calculate the expectation of the squared magnitude of the reverberant part of the response as:

$$E\{|h_r(n)|^2\} = \sum_{n,m} e^{-k(n+m)} E\{\bar{R}(n)\bar{R}(m)\} e^{-j(n-m)\omega}. \quad (\text{A.9})$$

Notice that since  $\bar{R}(n)$  is white, the  $E\{\bar{R}(n)\bar{R}(m)\} = \sigma_r^2$  only when  $n = m$  and is zero otherwise. Therefore we can define  $\epsilon_r$  as the energy of the reverberant tail of the room impulse response and express:

$$\epsilon_r = \sum_n |h_r(n)|^2 = \sigma_r^2 \sum_n e^{-2kn} = \frac{\sigma_r^2}{1 - e^{-2k}}. \quad (\text{A.10})$$

Expressing the direct path energy as  $\epsilon_d = \alpha_d^2$ , the output energy is given as:

$$\epsilon_y = (\epsilon_d + \epsilon_r) \epsilon_x, \quad (\text{A.11})$$

where  $\epsilon_x = \sum_n |x(n)|^2$ .

Level equalization as used in our experiment involves normalizing each stimulus by the square root of its total energy such that the normalized stimulus is<sup>3</sup>:

$$\tilde{y}(n) = \frac{y(n)}{\sqrt{\epsilon_y}} = \tilde{h}(n) * \tilde{x}(n), \quad (\text{A.12})$$

where the  $\tilde{h}(n) = h(n)/\sqrt{(\epsilon_d + \epsilon_r)}$  is the normalized room impulse response and  $\tilde{x}(n) = x(n)/\sqrt{\epsilon_x}$  is the normalized input signal. Therefore, the normalized direct path component of the impulse response becomes:

<sup>3</sup> Notice that, although we use a single impulse response in the derivations, the results apply also to the binaural stimuli used in our study since only sources in the front direction are employed.

$$\tilde{h}_d(n) = \frac{\alpha_d}{\sqrt{\alpha_d^2 + \epsilon_r}} \delta(n). \quad (\text{A.13})$$

Let us now define  $\alpha_d(D)$  as the true level of the direct path for a source in acoustic free-field at a distance  $D$  and let us parameterize the ratio of direct and reverberant energies such that  $\rho(D) = |\alpha_d(D)|/\sqrt{\epsilon_r}$ . Notice that  $\rho(D)$  is proportional to  $\alpha_d(D)$  since the energy of the reverberant component,  $\epsilon_r$ , is independent of the source and receiver positions in the room. The level normalized direct path becomes:

$$\tilde{h}_d(n) = \tilde{\alpha}_d(D)\delta(n) = \frac{\rho(D)}{\sqrt{[\rho(D)]^2 + 1}} \delta(n). \quad (\text{A.14})$$

Attenuation due to distance follows the inverse-square law and the following relation holds for the sound pressure levels due a source positioned at  $D_1$  and the same source positioned at  $D_2$ :

$$\Delta L_{D_1,D_2} = 20\log_{10} \left| \frac{\alpha_d(D_1)}{\alpha_d(D_2)} \right| = 20\log_{10} \left| \frac{D_2}{D_1} \right| \text{ (dB)}. \quad (\text{A.15})$$

Note also that since the reverberant energy is constant:

$$\frac{\rho(D_1)}{\rho(D_2)} = \left| \frac{\alpha_d(D_1)}{\alpha_d(D_2)} \right| = \frac{D_2}{D_1}. \quad (\text{A.16})$$

For the level normalized signal, the following relation holds:

$$\begin{aligned} \tilde{\Delta} L_{D_1,D_2} &= 20\log_{10} \left| \frac{\tilde{\alpha}_d(D_1)}{\tilde{\alpha}_d(D_2)} \right| \text{ (dB)} \\ &= \Delta L_{D_1,D_2} + 10\log_{10} \left| \frac{[\rho(D_1)]^2 + 1}{[\rho(D_2)]^2 + 1} \right|. \end{aligned} \quad (\text{A.17})$$

We will now show that for the skyway that we made our BRIR measurements, the change in the direct path levels of level-equalized stimuli follow the inverse-square law within a small and bounded error.

The critical distance,  $D_c$  in a room is the distance at which the energies of direct and reverberant components are equal such that  $\rho(D_c) = \alpha_d(D_c)/\sqrt{\epsilon_r} = 1$ . The critical distance for a room with volume  $V$  and a reverberation time of  $T$ , and with a sound source having a directivity gain of  $\gamma$  is given as [3]:

$$D_c \approx 0.1 \left( \frac{\gamma V}{\pi T} \right)^{\frac{1}{2}}. \quad (\text{A.18})$$

Using (A.16) and the definition of critical distance, it is possible to specify  $\rho(D)$  for all distances such that:

$$\rho(D) = \frac{D_c}{D}. \quad (\text{A.19})$$

Substituting (A.19) in (A.17), the level of the direct path component with and without level normalization are given respectively as:

$$\begin{aligned} L_D &= L_{D_c} + \Delta L_{D,D_c}, \\ \tilde{L}_D &= L_{D_c} + \Delta L_{D,D_c} + 10\log_{10} \left( \left[ \frac{D_c}{D} \right]^2 + 1 \right) - 3.01, \end{aligned} \quad (\text{A.20})$$

where  $L_{D_c} = 20\log_{10} |\alpha_d(D_c)|$ .

Notice that  $\lim_{D \rightarrow \infty} \tilde{L}_D = L_D - 3.01$  dB. In other words, the error induced by level equalization on the direct part component is always less than 3.01 dB.

Substituting  $V \approx 76 \text{ m}^3$ ,  $T_{60} = 1.15 \text{ s}$  and assuming an average directivity gain of  $\gamma \approx 2$  for the Genelec 6010 loudspeaker we used, the critical distance for the skyway that we made our measurements can be estimated as  $D_c \approx 0.65 \text{ m}$ . In order to make this difference more comprehensible within the range of sound pressure levels employed in this article, let us now designate the direct path level at the critical distance,  $\alpha_d(D_c)$  as an arbitrary reference, such that  $L_{D_c} = 70$  dB (SPL). Figure A.1 shows the actual and normalized direct path levels which differ from actual levels by 1.28 dB at the closest to  $-3$  dB at the furthest emulated distances. These differences are small and the distance attenuation trend closely follows the attenuation trend for the non-equalized case, especially beyond the critical distance. Therefore, the employed level equalization does not eliminate relative level cues pertaining to the level of the direct path component. In other words, the intensities of direct path components of level equalized stimuli also substantially attenuate by 6 dB per doubling of distance. Since the reverberant energy is assumed to be constant in a room regardless of the measurement position, the relative changes in the  $D/R$  ratio will also closely follow the non-equalized stimuli, retaining the relevant relative cues in the stimuli therein.

**Cite this article as:** Demirkaplan Ö & Hacıhabiboğlu H. 2020. Effects of interpersonal familiarity on the auditory distance perception of level-equalized reverberant speech. Acta Acustica, 4, 26.