PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

Audio/visual interaction in the perception of sound source distance

Pavel ZAHORIK[1]

[1] University of Louisville, USA

ABSTRACT

The perception of sound source distance is known to exhibit systematic biases. In general, the distances of far sources are progressively underestimated, but near sources are overestimated. Such biases are not typically observed in vision, however. Under natural viewing conditions in which a variety of visual distance cues are available to the observer, perceived distance is highly accurate. Relatively little is known about how distance information from both auditory and visual modalities is combined in the perception of distance, however. This is surprising, given that audio/visual aspects of directional perception have been extensively studied, primarily in relation to the "ventriloquist effect". Here, two experiments on audio/visual distance perception are summarized. Both used virtual auditory space techniques to simulate reverberant sound field listening of a loudspeaker-produced broadband noise signal. The results from both experiments suggest that not only is perceived distance less accurate in the auditory modality than in vision, but it is also considerably less precise. A computational model was developed based on data from these two experiments. Predictions from the model offer explanations as to why visual information, when available, appears to dominate auditory information in the perception of distance.

Keywords: Sound Localization, Spatial Hearing, Multisensory

1. INTRODUCTION

Auditory-visual interaction has been extensively studied in directional space. In general, the visual system provides superior directional accuracy and resolution, and therefore dominates the perceived direction of sound-producing objects. The dominance is strong enough to produce illusory percepts, such as the well-known ventriloquist situation, where the sound is localized to a plausible visual target even though that target does not actually produce the sound. The ventriloquist's illusion can influence sound sources separated from visual targets by as much as 55 degrees (1), appears to be strengthened by temporal synchrony between auditory and visual targets (2), but is unaffected by either attention to the visual distracter or feedback provided to the participant (3). Cortical level mechanisms have been shown to underlie the illusion (4, 5), which, along with associated aftereffects, suggest a type of short-term plasticity of perceived auditory space mediated by visual input (6).

Much less is known regarding auditory-visual interaction in the distance dimension. Pioneering work by Gardner (7) has suggested an even stronger visual dominance, where sound in anechoic space is always localized in depth to the nearest plausible visual target. Termed the "proximity-image effect", Gardner (7) demonstrated complete visual dominance over a range of 9 m between the more distance sound source and the visual target. It is important to note, however, that the auditory distance information available to listeners in these experiments was impoverished due to the use of an anechoic environment. In this type of environment, reverberant sound energy is effectively removed, which in turn removes an important acoustic cue to source distance, namely the ratio of direct to reverberant sound energy (8).

Given these facts, Mershon et al. (9) examined the proximity-image effect under more natural, semi-reverberant acoustical conditions, reasoning that the proximity-image effect may be due largely to auditory distance localization inaccuracies resulting from the anechoic conditions of Gardner's experiments. To test this reasoning, they visually presented observers with a realistic looking "dummy" loudspeaker in a semi-reverberant room and then played long duration (5-s) noise signals from a loudspeaker occluded from the observer's view either closer or farther away than the dummy loudspeaker. Ninety percent of the observers (n = 441) reported that the sound stimulus appeared to

[1] pavel.zahorik@louisville.edu

originate from the position of the dummy loudspeaker. Comparing this result to an anechoic condition in which 94% of observers (n = 96) reported that the noise stimulus appeared to originate from the dummy loudspeaker, it was concluded that the proximity-image effect operates with nearly the same strength in reverberant conditions as it does in anechoic conditions. It is also interesting to note that although the proximity-image effect was found to be the strongest when the dummy loudspeaker was closer than and the actual sound source, there was also some evidence of the effect in the reversed situation (dummy farther than actual sound source). Thus, there is evidence of a somewhat more general form of visual "capture" of auditory sources in the distance dimension, perhaps related to the angular direction capture reported in studies of ventriloquism effects.

Computational modeling efforts by Alais and Burr (10) have fundamentally changed the way the ventriloquist illusion is view conceptually. Previous to their innovative work, the illusion was viewed as a "winner take all" example of visual encoding of spatial information. Their modeling efforts for the ventriloquist illusion instead suggest a probabilistic view, where under most circumstances, the visual encoding of space is simply more reliable. Alais and Burr (10) confirm this hypothesis by demonstrating that auditory directional encoding can become dominant when directional information from vision is intentionally made unreliable. An additional and important prediction from Alais and Burr's (10) model is that the precision with which objects are localized in space is always better with multimodal input (auditory + visual) than with unimodal input (visual alone or auditory alone).

Mendonça et al. (11) apply this type of probabilistic explanation to visual capture in the distance dimension by evaluating a number of different probabilistic models. Although in general, this approach can explain situations both where visual capture in distance is (7, 9, 12) and is not observed (13, 14), the models evaluated by Mendonça et al. (11) do not incorporate a fundamental aspect of perceived auditory space: that perceived distance is non-linearly related to physical distance. Best evidence suggests that perceived distance is instead logarithmically related to physical sound source distance (see 8 for a meta-analysis of the auditory distance perception literature).

The purpose of this study is to extend probabilistic modeling of visual capture in distance by including consideration of the logarithmic relationship between perceived distance and physical distance, particularly in the auditory modality. Figure 1 shows a conceptualization of this space, where auditory distance percepts are less precise than visual distance percepts, and they systematically underestimate physical distance. A probabilistic model based on this conceptualization is then used to predict results from a psychophysical experiment in which participants judge whether or not the distance of a virtual sound source matches a visual target. Absolute distance estimates to both visual and (virtual) auditory targets were also collected, and used to model the accuracy of auditory and visual distance percepts (e.g. distribution means in Figure 1). Estimates of auditory and visual distance precision (e.g. distribution variances in Figure 1) were taken from data reported by Anderson and Zahorik (15). Because this work has already been published, written summary will not be provided here.

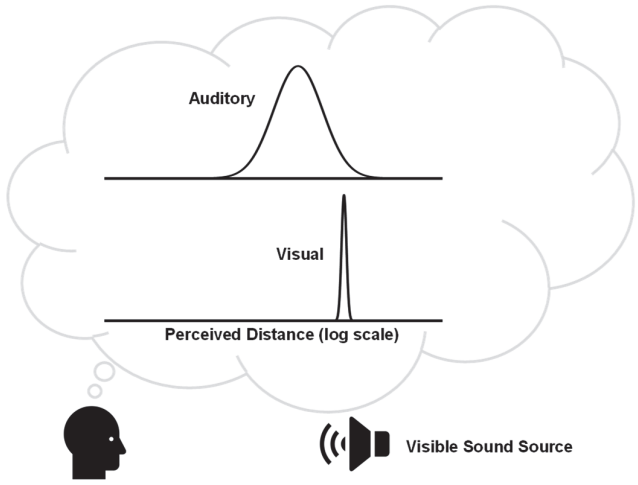

Figure 1 - Conceptual framework showing probabilistic distributions of perceived auditory and visual target distances. Note that both are represented on a logarithmic distance axis, and that perceived auditory distance both underestimates the physical distance to the sound source and is less precise. These aspects of the framework were motivated by results previous data sets (15, 16).

2. METHODS

2.1 Subjects

Eleven volunteers (4 male, 7 female; age range 18.1 – 19.4 years) participated in the psychophysical experiment. All had self-reported normal hearing and normal vision. All procedures involving human subjects were approved by the University of Louisville and University of California – Santa Barbara Institutional Review Boards.

2.2 Testing Environment

The experiment was conducted in a large office space ($7.7 \times 4.2 \times 2.7$ m^3) with carpeted floor, painted gypsum board walls, and drop acoustical tile ceiling. The room had an average background noise level of 31 dBA, and a broadband reverberation time ($RT60$) of approximately 0.6 s. The room was illuminated by ceiling-mounted fluorescent lighting (approximately 500 lux) typically used in office spaces. The listener was seated at one end of the room, approximately 1 m in front of the rear wall, and 2.1 m from the side walls. The experiment required two measurement phases in the test environment. The first phase measured the acoustical responses for various sound source distances at the ears of a single listener. These measurements were used to construct a virtual auditory space (VAS) used for subsequent phase two testing. Details of the VAS procedure are described in the next section. The second phase of the experiment measured listener's judgements of auditory/visual distance and coincidence in distance, using a real visual target (the measurement loudspeaker), and VAS to plausibly reproduce auditory targets independent of visual target location.

2.3 Virtual Auditory Space (VAS) Technique

A VAS technique, fundamentally similar to that described by Zahorik (16), was used to present virtual sounds over headphones at distances ranging from 1 to 5 m directly in from of the listener at ear height. To construct the VAS, seventeen binaural room impulse responses (BRIRs) were measured in the testing environment from distances ranging from 1 to 5 m in 0.25 m steps. The sound source was a small, full-range loudspeaker (Micro-spot, Galaxy Audio) placed on a stand at ear level (134 cm above the floor) powered by a high-quality amplifier (D-75, Crown). Miniature electret microphones (Sennheiser KE4-211-2) were placed in the ear canals (blocked-meatus configuration) of a single participant (the author), who did not participate in subsequent psychophysical testing. Maximum-length sequence (MLS) system identification techniques (17) were used to measure and derive the BRIRs. The measurement period was 32767 samples at a sampling rate of 44.1 kHz. To improve the signal-to-noise ratio of the measurements, 20 measurement periods were averaged for each distance. Post-averaging, the poorest measurement signal-to-noise ratio was 48 dB (C-weighted), which occurred at 5 m. The MLS measurement technique was implemented in Matlab (Mathworks, Inc.) using a high-quality digital audio interface (CardDeluxe, Digital Audio Labs). In order to equalize for the response of the headphones (Sennheiser HD 410 SL) used in VAS, the impulse responses of the left and right headphones were also measured using similar MLS techniques.

The source signal was a brief sample of broadband Gaussian noise, 100 ms in duration (1 ms rise/fall cosine gate). Independent samples were drawn for each stimulus presentation. No loudspeaker equalization was implemented, so the spectrum of the source signal was shaped by the loudspeaker response characteristics, which limited the bandwidth to between 150 Hz and 18 kHz. All auditory signal processing implemented using MATLAB software (Mathworks Inc., Natick, MA).

2.4 Visual Stimuli

The visual stimulus was the measurement loudspeaker, viewed binocularly, and placed at distances of either 1.5, 3, or 4.5 m.

2.5 Procedure

The experiment consisted of three psychophysical measurement phases: Yes/No judgements of Auditory/Visual target coincidence, absolute judgements of virtual auditory target distance, and absolute judgements of visual target distance. All participants completed all three phases of the experiment, in the order listed. Participants were not provided with any response feedback in any phase of the experiment.

Coincidence judgements. Participants were presented with a visual target and a virtual auditory target from either the same or different distances, and were instructed to respond whether the

perception was of matching (coincident) auditory and visual target distance, or not. Participants were told that the sound could actually originate from the loudspeaker visual target, or be a virtual sound that was produced by the headphones, even though in actuality all sounds were virtual. This was done to avoid potential biases that could result if participants knew that physical coincidence was impossible. This was also the rationale for testing coincidence prior to absolute judgements, where exposure to virtual sounds alone was known to participants. To facilitate good registration between auditory and visual targets, head orientation was monitored using a laser pointer. The pointer was mounted to the headband of the headphones and pointing straight ahead. Participants were instructed to keep the laser pointer aimed at a 2-cm circular target affixed to the front of the loudspeaker visual target. Compliance with this instruction ensured orientation remained fixed within one degree. Head orientation compliance was monitored by an experimenter, who controlled the initiation of each trial and entered the participant's coincidence responses on a keypad. Testing was conducted blocked by visual target distance. For each visual target distance, all 17 virtual auditory target distances were tested 20 times, in randomized order. Visual target block order was also randomized. Participants completed this portion of the experiment in approximately 30 minutes.

Auditory distance judgments. Participants made absolute judgements of virtual auditory target distances in the absence of any plausible visual targets. Virtual auditory target distances ranged from 1 to 5 m in 0.5 m steps. Head movement was monitored using procedures identical to those used for coincidence judgements, except that the circular target to which the participant was instructed to aim the pointer was mounted on the back wall of the testing room. Participants could use distance units with which they had the most familiarity (e.g. either feet/inches, or meters/centimeters), and were instructed to be as precise as possible with their estimates. A single distance judgement from each distance was recorded. Presentation order of distances was randomized. Participants completed this portion of the experiment in approximately 5 minutes.

Visual distance judgments. Judgments of absolute distance were collected for the three visual target distances in the absence of auditory targets. Procedures were identical to those used for auditory distance judgements, except that the head orientation target was fixed to the loudspeaker visual target. Participants completed this portion of the experiment in approximately 2 minutes.

3. RESULTS AND DISCUSSION

In order to avoid duplication with an upcoming publication, results will be described here only in summary form. Overall, coincidence judgement results were consistent with the proximity-image effect (7, 12): Nearby visual targets more effectively captured faraway sound sources than vice versa. Further, because this effect was demonstrated in a reverberant sound field, it is consistent with the observation by Mershon et al. (9) that the effect is not specific to anechoic space where auditory distance information is impoverished. This effect was well-predicted by a probabilistic model of auditory/visual integration motivated by the conceptual framework shown in Fig. 1. that incorporates a logarithmic perceptual space in which auditory distance percepts are more biased and more variable than visual distance percepts.

4. CONCLUSIONS

A key component to successfully explaining the proximity-image effect specially, and visual capture in the distance dimension more generally, appears to be the highly non-linear aspects of perceived auditory space. Predictions of audio/visual coincidence judgements in the distance dimension were found to be most accurate for a probabilistic model that combines auditory and visual information assuming a logarithmic perceptual space.

ACKNOWLEDGEMENTS

This work was supported in part by the University of Louisville and grants from the NIH (F32EY07010 & R21EY023767) and AFOSR/KYDEPSCoR (FA9550-08-1-0234).

REFERENCES

1. Thurlow WR, Jack CE. Certain determinants of the "ventriloquism effect." Perceptual & Motor Skills. 1973;36(3, Pt. 2):1171-84.
2. Jack CE, Thurlow WR. Effects of degree of visual association and angle of displacement on the

"ventriloquism" effect. Perceptual & Motor Skills. 1973;37(3):967-79.
3. Bertelson P, Vroomen J, De Gelder B, Driver J. The ventriloquist effect does not depend on the direction of deliberate visual attention. Perception & Psychophysics. 2000;62(2):321-32.
4. Bonath B, Noesselt T, Martinez A, Mishra J, Schwiecker K, Heinze HJ, et al. Neural basis of the ventriloquist illusion. Current Biology. 2007;17(19):1697-703.
5. Callan A, Callan D, Ando H. An fMRI study of the ventriloquism effect. Cerebral Cortex. 2015;25(11):4248-58.
6. Recanzone GH. Rapidly induced auditory plasticity: the ventriloquism aftereffect. Proceedings of the National Academy of Sciences U S A. 1998;95(3):869-75.
7. Gardner MB. Proximity image effect in sound localization. Journal of the Acoustical Society of America. 1968;43(1):163.
8. Zahorik P, Brungart DS, Bronkhorst AW. Auditory distance perception in humans: A summary of past and present research. Acta Acustica united with Acustica. 2005;91:409-20.
9. Mershon DH, Desaulniers DH, Amerson TLJ, Kiefer SA. Visual capture in auditory distance perception: Proximity image effect reconsidered. Journal of Auditory Research. 1980;20:129-36.
10. Alais D, Burr D. The ventriloquist effect results from near-optimal bimodal integration. Current Biology. 2004;14(3):257-62.
11. Mendonça C, Mandelli P, Pulkki V. Modeling the perception of audiovisual distance: Bayesian causal inference and other models. PLoS One. 2016;11(12):e0165391.
12. Gardner MB. Distance estimation of 0 degrees or apparent 0 degree-oriented speech signals in anechoic space. Journal of the Acoustical Society of America. 1969;45(1):47-53.
13. Calcagno ER, Abregu EL, Eguia MC, Vergara R. The role of vision in auditory distance perception. Perception. 2012;41(2):175-92.
14. Zahorik P. Estimating sound source distance with and without vision. Optometry and Vision Science. 2001;78(5):270-5.
15. Anderson PW, Zahorik P. Auditory/visual distance estimation: accuracy and variability. Frontiers in Psychology. 2014;5:1097.
16. Zahorik P. Assessing auditory distance perception using virtual acoustics. Journal of the Acoustical Society of America. 2002;111(4):1832-46.
17. Rife DD, Vanderkooy J. Transfer-function measurement with maximum-length sequences. Journal of the Audio Engineering Society. 1989;37(6):419-44.