

Knowledge about typical source output influences perceived auditory distance^{a)} (L)

John W. Philbeck^{b)} and Donald H. Mershon

Department of Psychology, North Carolina State University, Raleigh, North Carolina 27695

(Received 17 July 2001; accepted for publication 26 February 2002)

Vocal effort is known to influence the judged distance of speech sound sources. The present research examined whether this influence is due to long-term experience gained prior to the experiment versus short-term experience gained from exposure to speech stimuli earlier in the same experiment. Speech recordings were presented to 192 blindfolded listeners at three levels of vocal output. Even upon the first presentation, shouting voices were reported as appearing farthest, whispered voices closest. This suggests that auditory distance perception can be affected by past experience in a way that does not require explicit comparisons between individual stimuli. © 2002 Acoustical Society of America. [DOI: 10.1121/1.1471899]

PACS numbers: 43.66.Qp, 43.66.Lj, 43.71.Bp [LRB]

I. INTRODUCTION

Egocentric distance is the distance between an observer and a point in space; perceived egocentric distance in the auditory domain is the apparent distance between a listener and a sound source. Stimulus information that influences this perception includes the intensity of the sound reaching a listener's ears and the ratio of direct to reflected sound in a given environment (Bronkhorst and Houtgast, 1999; Mershon and King, 1975; Zahorik, 1998). In addition to these stimulus variables, a listener might also determine the source distance of familiar sounds by comparing the sound pressure level at the ears with some internal estimate of the probable output power of the sound source. For example, given a very faint proximal stimulus that one identifies as a fire engine siren, one might perceive the source to be far away, because sirens usually have high output power. Familiarity with a sound source can encompass many different kinds of information, but here we will define "source familiarity" more specifically to mean the stored knowledge upon which one might base such estimates of output power. As yet, only a few studies have systematically investigated the influence of sound source familiarity on the perception of auditory distance (Brungart and Scott, 2001; Gardner, 1969). This research has focused on speech sounds, and the results clearly show that when other stimulus factors are held approximately constant, estimates of the source distance of speech sounds are modulated by the production level used in generating the speech;¹ specifically, listeners indicate the source distance of whispered speech to be nearer than that of shouted speech.

A critical question remains unanswered by the foregoing research, however. The previous studies tested a relatively small number of listeners and averaged across multiple judg-

ments per condition per listener when analyzing the data. This being the case, one cannot determine whether the source familiarity effects are due to long-term experience with the typical production level of speech or are instead due to repeated exposures to particular speech sounds within the immediate experimental setting. If the latter is true, the perceived source distance of the *first* speech sample in an experiment could be determined primarily by reverberation, absolute intensity, or some other kind of distance information that does not depend upon source familiarity (Mershon and King, 1975). Listeners might then base their distance estimates in subsequent presentations on *changes* in the perceived production level of the speech samples, relative to their initial estimate. In this way, previous reports of source familiarity effects (Brungart and Scott, 2001; Gardner, 1969) may be due to comparisons between stimuli within the immediate experimental context and have nothing to do with long-term experience with speech sounds. This letter describes a control experiment designed to rule out this possibility. To prevent individuals from comparing stimuli across trials, one must analyze the data obtained from the very first stimulus presentation. If *long-term* knowledge about the typical source output of speech contributes to the perceptual localization of the sound source relative to the listener, production level should affect source distance judgments even upon the first stimulus presentation.

II. METHOD

A. Testing environment

Testing was conducted in a carpeted room, 7.3×7.3×3.7 m, with an average reverberation time (T_{60} , the time required for a sound to decay by 60 dB) of approximately 0.3 s across the frequencies of interest. The stimuli were presented by a Polk Audio (Model 5) loudspeaker system 2.5 m from the listener's head, positioned approximately at ear level in the median plane. The listener stood in front of a sound-absorbing wedge which reduced reflections from the wall behind the listener. The straight line between the listener

^{a)}Portions of this work were presented at the 32nd annual meeting of the Psychonomic Society, 22–24 November 1991, San Francisco, CA.

^{b)}Current address: Department of Psychology, The George Washington University, 2125 G Street, NW, Washington, DC 20052. Electronic mail: philbeck@gwu.edu

and loudspeaker was parallel to two walls, but slightly offset from the room's center-line. Thirty-six overhead loudspeakers (12.7 cm diameter) created a diffuse wideband masking noise to hide noise intrusions from outside the laboratory. The sound level of this noise at the listener's ears was 48 dBA. Previous work involving this room in a similar configuration demonstrated that sufficient reverberation remains to generate some modulation in distance estimates for stimuli consisting of white noise bursts (Mershon *et al.*, 1989). Median distance estimates increased by just over a meter for a range of source distances between 0.75 and 6 m; at 3 m (nearly the same source distance as in the present study), the median response was approximately 1 m. Because the source distance did not vary in the current experiment, reverberation information signaled the same source distance for each production level. Under these conditions, distance judgments could be biased toward the source distance given by reverberation (perhaps near 1 m, based on the findings of Mershon *et al.*, 1989).

B. Generation and presentation of stimuli

The sound stimuli were recorded in the testing room described above. One male and one female talker were recorded speaking the phrase "How far away from you does my voice seem?" Each talker provided a sample of the phrase using a whisper, a shout, and a normal conversational level² with the microphone positioned approximately 30 cm in front of the talker's mouth. For whispered recordings, the talkers whispered as if communicating with someone at the distance of the microphone; for conversational recordings, they used a voice appropriate for communicating with someone just beyond arm's reach; for shouted recordings, they attempted to shout as loudly as possible. Speech samples were digitized at 44.1 kHz with 16-bit resolution; during playback, the samples were amplified by a Crown DL-2 pre-amplifier and Crown PS-200 amplifier before being sent to the loudspeaker.

A Rion NA-61 Impulse Precision Sound Level Meter (with an NA-2X third-octave filter set) was used to obtain average and peak sound levels. Spectrographic analysis showed that, not surprisingly, the male voice included lower frequency components than did the female voice. The different production levels also showed clear variation. The whispered stimuli generally lacked the very low frequency energy associated with voicing and was dominated by energy in the middle and upper frequency ranges. The shouts tended to be dominated by lower frequency components associated with voicing of vowels. Conversational speech fell somewhere between these extremes.

Ideally, one would want the sound level of all stimuli to be equal in order to eliminate this as a controlling factor. We adopted the conservative approach of setting both average and peak levels of the shouting voice to be slightly *higher* than the corresponding values for the whispered and conversational voices (see Table I). This ensured that, whatever the contribution of sound level, it should have worked *against* the expected perception of a distant shout.

TABLE I. Average and peak sound levels of the three different speech samples by the male and female talkers, as presented on playback to listeners and measured at the position of the listener's head. All values given in sound levels (dBA). These stimuli were heard against a background of wide band masking noise presented at 46 dBA.

	Male voice		Female voice	
	Average	Peak	Average	Peak
Shouted	72	78	72	82
Conversational	66	74	67	77
Whispered	66	74	67	76

C. Listeners

A total of 192 listeners (96 men, 96 women) participated in this experiment for course credit. All reported normal hearing in both ears. None had previously seen the laboratory.

D. Design

Listeners were randomly divided into six groups of 32 listeners. The groups were distinguished by which talker and which production level was heard on the first presentation. Following the initial presentation of one of the six possible stimuli, each listener was then presented with the other two production levels, using the same voice (male or female) heard on the first presentation. Finally, the listener was presented again with the sample heard initially. Thus, each listener separately contributed an initial report for one of the samples, followed by additional reports for all three samples spoken in the same (male or female) voice. This design allowed for nonoverlapping analyses for first presentations and for presentations with an explicit preceding comparison stimulus. Each of the possible orderings of stimuli occurred equally often.

E. Procedure

Listeners were blindfolded before entering the testing room. They were never given prior exposure to speech from a distance within the testing environment, nor did they hear the stimulus voices before the actual stimulus presentation. The four stimuli were presented in sequence, with the listener verbally judging the source distance after each stimulus. The instructions strongly emphasized that reports should be based on the *apparent* source distance, as opposed to trying to estimate the objectively accurate distance. This methodological detail is known to enhance the influence of perceptual factors in determining the response over explicitly cognitive factors (for a review, see Carlson, 1977). The listeners presumably knew the testing environment was indoors, but the instructions were carefully worded to avoid suggesting any particular real or imagined size of the testing room. Neither vision of the workspace nor error feedback was provided until after the final response.

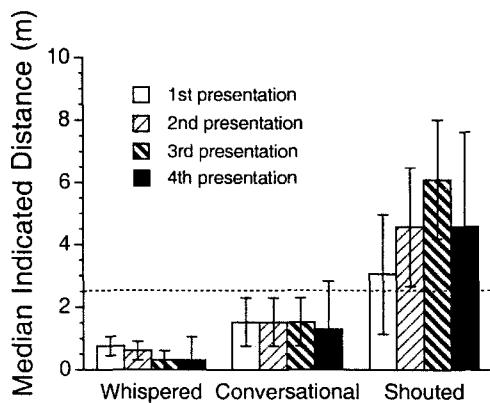


FIG. 1. Median indicated distances, showing data from presentations 1–4 for three production levels. Error bars represent ± 1 semi-interquartile range. The horizontal dashed line indicates the physical distance of the loudspeaker. One estimate of 402 m (first presentation, shouted stimulus) was omitted for this figure, but otherwise each bar represents $n = 64$.

F. Results

1. First presentation data

The median distance judgments for the four consecutive stimulus presentations are shown in Fig. 1. There is a clear increase in indicated distance across the three styles of speech even upon the first stimulus presentation. Median response values for whispered, conversational, and shouted speech were 0.76, 1.52, and 3.05 m, respectively, when these stimuli occurred first in the running order. The physical source distance was 2.5 m in all cases; the general inaccuracy of the responses is very likely a consequence of the limited availability of stimulus information specifying the source distance (see Mershon *et al.*, 1989). Of greater interest than the general pattern of inaccuracy, however, is the pronounced and systematic changes in distance judgments across the three production levels. Some skewing was apparent in the data, so an analysis of variance (ANOVA) using the ranks of the indicated distance values was performed. The rank transformation discards the original estimates and retains only the ordinal relations among them; the result is a test that, by making fewer assumptions about the data, is less sensitive (i.e., is *less* likely to detect differences between groups than before the transformation). This conservative analysis showed that the production level variable was significant ($F_{2,180} = 59.17$, $p < 0.0001$), with no other significant main effects or interactions.

2. Later presentations

When the listeners had the opportunity to make comparisons across the different stimulus presentations, the reports of distance for the whispered and shouted voices became more different. An ANOVA was performed on the ranked values of these repeated-measures data (presentations 2–4). In addition to a main effect of production level ($F_{2,360} = 790.69$, $p < 0.0001$), there were significant main effects of the sex of the listener ($F_{1,180} = 11.17$, $p = 0.001$) and of the sex of the voice used as a stimulus ($F_{1,180} = 4.22$, $p = 0.0415$). There was also a significant interaction of production level and the presentation order, i.e., which production level was presented on the initial trial ($F_{4,360} = 4.71$,

$p = 0.001$). Differences between reports from male and female listeners may represent a genuinely perceptual difference or, more probably, some difference in how each group assigned numbers to a common perceptual experience. The difference associated with the sex of the talker may be related to differences in the typical output power of male shouts relative to female shouts. At present, the effects of individual differences on perceived auditory distance are poorly understood (although see Brungart and Scott, 2001, for one analysis).

III. CONCLUSIONS

There are two main conclusions. First, listeners clearly report whispers, conversational speech, and shouts at systematically different distances, even upon initial presentations and under conditions in which prior conceptions about the possible source locations are minimized. Analysis of the first presentation data of nearly 200 listeners firmly establishes that the effects of source familiarity are the result of long-term experience with speech sounds, rather than comparisons between speech stimuli encountered within the immediate experimental context. Variations in production level from shouting to whispering were associated with changes in distance judgments by as much as a factor of 4 (medians: 3.05 m vs 0.76 m). By demonstrating that source familiarity affects egocentric distance estimates even when comparisons with other experimental stimuli have been prevented, we have shown that source familiarity provides absolute distance information (Gogel, 1968; Mershon and Bowers, 1979). Second, the results shed light on the time course of the accretion of information across multiple stimulus presentations. Specifically, the median distance estimates changed systematically over a very short time scale, on the order of only a few trials. Presumably, after the listeners responded to the initial stimulus presentation, their distance estimates were influenced by a combination of two kinds of source familiarity: (1) long-term experience with speech obtained prior to the experiment, and (2) short-term experience with speech stimuli presented earlier in the experiment. Although we did not attempt to determine the relative contribution of these two sources in the trials following the initial stimulus presentation, it is clear that the effect of source familiarity was heightened when the two kinds of familiarity were available in combination (Fig. 1). The very rapid change in distance estimates upon repeated exposure to a single voice may explain why previous work has found virtually no effect for prior exposure to a talker's voice (Brungart and Scott, 2001); if such changes become attenuated very rapidly and reach a steady state, the effect will likely become more and more diluted upon additional stimulus presentations.

In studies that use direct verbal distance estimates, it is difficult to dissociate genuine perceptual influences from more abstract cognitive influences (e.g., reasoning). Even if the verbal estimates reflect a composite of perceptual and cognitive factors, however, this composite signal behaves in a very stable and predictable manner with changes in production level. Taken together, our results and those of previous researchers (Brungart and Scott, 2001; Gardner, 1969) indicate that source familiarity is indeed a potent determinant

of perceived auditory distance, operating under a variety of conditions—across many listeners and talkers, inside and outside the laboratory, with and without vision, and using both live and prerecorded speech stimuli. These factors suggest that source familiarity can be exploited successfully to convey distance information in both real and virtual environments. The relatively large perceived distances that this information is able to generate suggests that it can contribute to the guidance of human navigation on the basis of auditory information. These results also show promise for applications concerning the design of auditory displays to minimize attentional demands in high-workload situations.

ACKNOWLEDGMENTS

The authors thank William Franklin for use of the speech spectrograph, Elliott Inman and Mary Catherine Bunn for their voice work, Pat Cox for technical assistance, and David Clarke, Robert Remez, and Pavel Zahorik for helpful comments on earlier drafts.

¹“Production level” denotes the sound pressure level of a talker’s voice measured from a fixed distance near his or her head (Brungart and Scott, 2001). A related term is “vocal effort,” which refers to the amount of stress or force imparted to a vocal utterance (Traunmüller and Eriksson, 2000). Stimulus information signaling vocal effort may be used to estimate the production level of a particular utterance. Vocal effort and production level are tightly linked, but in certain portions of the vocal effort continuum, the correlation is not perfect. An unvoiced “stage whisper,” for example, might result in a higher production level than a quietly voiced utterance, even though the whisper might entail less apparent vocal effort than the voiced

speech. The linkage between vocal effort and production level is sufficiently close, however, that production level provides a useful characterization of vocal effort (Brungart and Scott, 2001).

²The output power of the speech samples was not directly measured, but other research (Traunmüller and Eriksson, 2000) has shown that the sound-pressure level of whispers is typically about 40 dB or less, relative to an arbitrary reference, with conversational-level speech registering at around 60 dB and shouts at 85 dB or more.

- Bronkhorst, A. W., and Houtgast, T. (1999). “Auditory distance perception in rooms,” *Nature (London)* **397**, 517–520.
- Brungart, D. S., and Scott, K. R. (2001). “The effects of production and presentation level on the auditory distance perception of speech,” *J. Acoust. Soc. Am.* **110**, 425–440.
- Carlson, V. R. (1977). “Instructions and perceptual constancy judgments,” in *Stability and Constancy in Visual Perception: Mechanisms and Processes*, edited by W. Epstein (Wiley, New York), pp. 217–254.
- Gardner, M. B. (1969). “Distance estimation of 0° or apparent 0°-oriented speech signals in anechoic space,” *J. Acoust. Soc. Am.* **45**, 47–53.
- Gogel, W. C. (1968). “The measurement of perceived size and distance,” in *Contributions to Sensory Physiology, Vol. III*, edited by W. D. Neff (Academic, New York), pp. 125–148.
- Mershon, D. H., and Bowers, J. N. (1979). “Absolute and relative cues for the auditory perception of egocentric distance,” *Perception* **8**, 311–322.
- Mershon, D. H., and King, L. E. (1975). “Intensity and reverberation as factors in the auditory perception of egocentric distance,” *Percept. Psychophys.* **18**, 409–415.
- Mershon, D. H., Ballenger, W. L., Little, A. D., McMurtry, P. L., and Buchanan, J. L. (1989). “Effects of room reflectance and background noise on perceived auditory distance,” *Perception* **18**, 403–416.
- Traunmüller, H., and Eriksson, A. (2000). “Acoustic effects of variation in vocal effort by men, women, and children,” *J. Acoust. Soc. Am.* **107**, 3438–3451.
- Zahorik, P. (1998). “Experiments in Auditory Distance Perception,” Ph.D. thesis, University of Wisconsin—Madison.