Familarity of Sounds as a Cue of Auditory Distance Perception

*Yoon-Ki Min

Abstract

The present research examined the contribution of sounds' familiarity to auditory distance perception, while attempting to control the influences of unavoidable physical characteristics among sounds. Different vocal "styles" ("shouts," "whispers" and "a normal conversation") of man and woman were recorded digitally and presented from a stationary loudspeaker to blindfolded listeners in a semi anechoic chamber. Playback levels were adjusted to remove extraneous sound level cues. The results showed that the shouting voice was judged as appearing farthest, the whispering voice closest, and the conversational voice was intermediate. The findings suggested that the perception of auditory distance may be affected by past experience (or familiarity).

I. Introduction

It has been demonstrated that listeners have the ability to estimate distance of sound sources with considerable accuracy. Perceived auditory distance of a sound source is the apparent distance between a listener and a sound source. Several characteristics, including the relationship of reverberant sound to direct sound and changes in the sound level or spectral content, of a sound source conjointly provide information to the listener for the sense of perceived distance[1-3]. Additionally, there have been many suggestions that familiarity with a sound may influence the perceived distance. The present study focused on the familiar sound and the perceived distance of such sound.

With regard to sound level, judgments of distance systematically increase as the level at the listener's ear decreases with changes in physical distance[4, 5]. Also the existence of direct and reflected sound energy, which occurs in most natural acoustic environments, facilitates the perception of sound-source distance. That is, the ratio of direct to reverberant sound decreases with distance[5, 6, 8]. There is another distance cue, spectral content cue. Sounds lacking high frequency components usually seem farther away than sounds containing high frequencies[1, 7, 8].

Finally, there is the possibility that familiarity (or past experience) with sounds may affect the distance of a source[9-11]. That is, the prior knowledge concerning the appropriate characteristics of the given sound be sufficient to create a perception of distance even on an initial presentation under new, experimental conditions. If familiarity does contribute to current perceptual experience by providing a cue to distance, then one might reasonably expect that it would do so by providing an absolute cue. That is, the prior knowledge concerning the appropriate characteristics of the given sound, combined with the stimulus characteristics of the present stimulus (e.g., its sound level at the ears, its spectral content, etc.), should be sufficient to create a perception of distance even on an initial presentation under new, experimental conditions. Note that, although the familiar sound cue depends upon registration of stimulus variables such as sound level and spectral content, the unique aspect is the information remembered from past experience. The characteristics of the familiarity cue are the subject of the present experiments.

* Department of Consciousness Science, Korea Research Institute of Jungshin Science/ Department of Design, Sejong University

Manuscript Received : July 12, 2000.

II. Present Studies and Goal

Since samples of human speech provide the most convenient familiar sounds, it is worthwhile to consider some of the ways in which such stimuli vary, even for statements having the same verbal content. First, one might compare male and female voices under conditions of normal conversational speech. Second, one can examine the changes in output which result from using different "styles," such as whispered or shouted speech. Perhaps the most obvious difference between male and female conversational voices is that the fundamental frequency is lower for male voices (typically 80-240Hz, compared with 140-500Hz for female voices)[14]. The male voice also tends to be capable of greater output power, although during normal speaking, the ranges of male and female voices do not differ much in level (73-89dB for men vs 73-80dB for women)[15]. Both of these differences

are associated with variations in the physiologic construction of the vocal apparatus, perhaps being accounted for solely by a difference in the length of the membranous vocal folds[16].

Although there is still an active debate concerning whether or not listeners perceive human speech in terms of the articulatory processes necessary to produce it[17], there is greater agreement that listeners are quite capable of perceiving the relative effort required to produce different speech sounds. This is true even for different vowel sounds made during normal conversational speech. Vowels spoken with equal emphasis may be judged equally loud despite a difference in their physical strength; intentional increases in speaking emphasis (to equate the same vowel sounds physically) result in judgments of unequal loudness[15]. Given this ability to interpret small changes in the effort exerted by a speaker, it seems certain that differences in production effort between shouts, whispers and conversational speech should be easily heard by adult listeners.

There is another reason for the current attempt to re-examine the familiar sound cue. Despite everyday experiences which suggest that we can perceive sounds to be at many distances, from very nearby to extremely far (perhaps even across a wide field), it has proven difficult in many laboratory situations to create the stable percept of a remote sound source. Even more notable, it has proven difficult to create the perception of a sound which reliably seems to originate from farther away than its actual source. The occasional report of very large auditory distances often involve stimuli presented in special environments such as a cathedral[18]. In the laboratory, with live presentations of sound from a loudspeaker, it has proven difficult to create consistent errors of over-perception. This should not taken as a statement that human simply perform too well under controlled conditions; it is not difficult at all to create the impression that a sound is closer than its true source.

The possibility that recorded familiar sounds might be able to create the appropriate sorts of perceptual error (and do so with some consistency) was another motive for conducting the present studies. Note that for this purpose it is critical that the reports of distance reflect genuinely perceptual changes and not merely cognitive adjustments of one's responses.

Gardner (1969)[9] examined the effects of different styles of speech on distance judgments, using presentations from live human, as well as recorded speech stimuli delivered from loudspeakers. The results indicated that shouted speech tended to produce distance estimates which were greater than those for conversational speech. Whispered speech led to the under-estimation of distance. Unfortunately, the only explicit data presented on this effect came from conditions in which a live person was used as the source of the sample. This situation might have obvious problems of experimental control. Moreover, reports of distance were obtained under instructions which emphasized accuracy (listeners had to identify which of 4 or 5 numbered accuracy), rather than appearance. This type of procedure has been criticized for unduly restricting listener's responses[5].

Finally, all of Gardner's studies were conducted in an anechoic space, with listeners who heard a variety of sounds during the same session. Such anechoic conditions would have created a strong cue for a very close distance (a factor that may have conflicted with the identification task used for reporting distances).

Therefore, the present experiments were carried out to verify that differences in speech style would still produce systematic variations in perceived distance. For this goal, the study used (1) an open-ended verbal response task, rather than limiting responses to a few specified alternatives, and (2) an acoustic environment with enough reverberation to eliminate the strong influence of anechoic presentations on perceived distance. The study also compared reports for initial presentations of the different speech stimuli to independent groups of listeners, as a way of determining the absolute-relative status of the familiar sound cue. To begin an evaluation of whether the effect of speech style is genuinely perceptual, rather than a conscious adjustment of verbal responses to fit listeners' expectations that shouts "should be" far and whispers "should be" close, the present instructions emphasized reporting the apparent distance of each sound, rather than accuracy of judgment. In addition, the study attempted to control the direct influences of sound level and spectral cues on the perception of auditory distance.

III. Methods

3.1. Environment and Response

All experiments were conducted in an enclosed, windowless test room consisting of a 7.3 x 7.3 x 3.6m (l x w x h) space covered with sound-absorbing panels to reduce reflections. The T60 reverberation time was approximately 0.36 s for frequencies between 0.5 and 8.0 kHz. Sonex® 4 inch acoustical foam (1.2 x 1.2m) was attached to the wall approximately 1.2m behind the observer. This panel was intended to eliminate early reflections which could otherwise have been produced by the wall behind the listener.

All presentations to listeners took place from a Polk

Audio (model 5) loudspeaker system, positioned 2.5m from the listener's head in the median plane. The straight line between the listener and the loudspeaker was parallel to two of the walls of the room.

For all presentations of the stimuli, blindfolded listeners reported the apparent distance to each sound in "feet, inches or in some combination of feet and inches." Listeners did not have an opportunity to view the laboratory until after completion of their last judgment. The experimenter always remained in the testing room with the listener, triggered the appropriate sounds from a remote keyboard and recorded the listener's responses. A continuous background of wide-band noise (with a sound level of 48dBA at the position of the listener's head) was presented from overhead speakers.

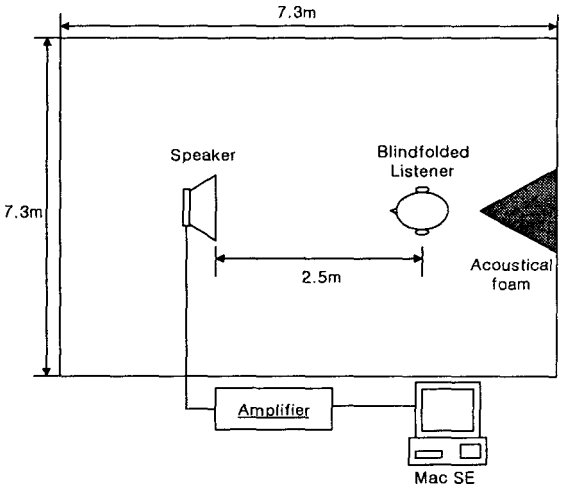

Figure 1. Experimental arrangement for presenting sounds.

3.2. Stimuli

The stimuli to be used as experimental sounds were recorded within the same acoustical environment in which they were later to be presented. Two volunteers (one male and one female) were recorded speaking the phrase "How far away from you does my voice seem?" Specifically, each speaker provided a sample of the phrase using a whisper, a shout and a normal conversation style. During recording the microphone was positioned approximately 30.5cm in front of the speaker's mouth.

Speech samples were digitized with 16-bit resolution by an Apple Macintosh SE computer with a Sound Accelerator® digital-processing board from digidesign. The digitized samples were stored as WAV files and all manipulations were accomplished using Sound Designer II software. For playback, samples were amplified by a Crown DL-2 preamplifier and Crown PS-200 amplifier before being sent to the loudspeaker.

Because of the internal variation within each speech sample, two measures of level were obtained. First, an "average" level was obtained by visual observation of the range of values shown on the sound level meter (Rion NA-61) during periods when the level remained roughly constant. Second, a peak level was measured by using the impulse setting and recording the highest value reached by the meter.

A third-octave spectral analyses showed that there were obvious differences among the samples and predictable differences among the samples. The male voice included lower frequency components than did the female voice. With regard to the styles of speech, the whispered voice generally lacked the very low frequency energy associated with voicing, but did contain energy in the higher ranges. The shouts, on the other hand, tended to be dominated by strong low frequency components associated with voicing of the vowel sounds. Conversational speech fell somewhere between these extremes. Each sample, of course, varied internally, due to the specific sounds required for pronounciation of the chosen phrase. The resulting levels are shown in Table 1.

Table 1. Average and peak Sound Levels of the Three Different Voice Samples by the Male and Female Speakers (All values Given in dBA).

	Male Voice		Female Voice	
Type of Voice	average	peak	average	peak
Shouted	73	77	71	82
Conversation	67	74	67	78
Whispered	66	74	67	75

The present experiment particularly interested in the shouted sample, because it is expected that the shouted voice was most likely to be over-perceived in distance. The experiment was, however, concerned that effects due to sound level or spectral differences not be mistaken for a familiarity effect. Unfortunately, equating diverse speech stimuli for some level (or for loudness) is not easy. Thus, is was unsure whether to equate some sort of average values or to equate peak levels. Perhaps the critical information, with respect to the use of sound level as a distance cue, was contained in the initial portion of each stimulus. With no certain guide, the experimenter chose to err on the side of caution by setting both "average" and peak levels of the shouting voice to be slightly higher than the corresponding values

for the whispered and conversational voices.

Leaving the sound level of the shout somewhat high on playback ensured that, whatever the contribution of sound level, it should have worked against the expected perception of a distant shout.

3.3. Listeners

A total of 192 college students (half men, half women) served as the listeners for this experiment. All reported that they had normal hearing in both ears. None of the listeners had previously seen the laboratory, nor had any listener served previously in related research.

3.4. Experimental Design

Six groups of 32 listeners each were given an initial presentation of one of six possible speech stimuli (gender x speech style). Following the initial presentation, each listener was then presented with the other two speech styles, using the same voice (male or female) heard on 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 voice. For all presentations of stimuli, blindfolded listeners reported the apparent distance to each sound in metric unit.

IV. Experimental Results

Two separate analyses were carried out: An analysis of the reports of perceived distance for the first (initial) presentation of a new voice using one of the three speech styles and an analysis of the distance reports for all three styles presented (later presentation) in the same voice (following the initial presentation). The skewness and the large variability of the perceived distance data is common for verbal judgments[13]. Thus, standard analytical techniques could not be applied to the original data sets. Instead, a rank-transform ANOVA[12] were employed. For this, all responses were transformed to rank cases. The new data set was then employed as the dependent variable in a standard ANOVA. Also, the median was used for showing the perceptual tendency.

4.1. First Presentation

Figure 1 shows the overall results of the initial presentations to the independent groups of listeners. There is a clear increase in the median reported perceived distance across the three styles of speech and a small over-perception of the distance to the shouting voice. A majority of both male and female listeners reported that the shout was perceived to be farther than the distance of the actual loudspeaker. A rank-transform ANOVA was performed because verbal reports of perceived distance are often skewed[13]. The main effect of "Speech Style" was significant ($F2,180=59.17$, $p<.001$). No other variables or interactions were significant.

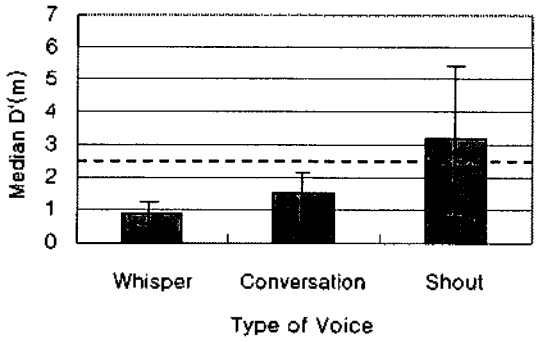

Figure 2. Median perceived distances (D') for the first presentation, using three different styles of voice. Error bars represent ±1 semi-interquartile range. The horizontal dashed line indicates the actual physical distance of the loudspeaker.

4.2. Later Presentation

The results from presentation 2-4 showed that the reports of distance for the whispered and shouted voices became more different, with comparisons available among the different styles (see Figure 3). The main effect attributable to speech style was significant ($F2,360=790.69$, $p<.001$). In addition, there were significant main effects of the sex of the listener ($F1,180=11.17$, $p<.001$) and of the sex of the voice used as a stimulus ($F1,180=4.22$, $p< .05$).

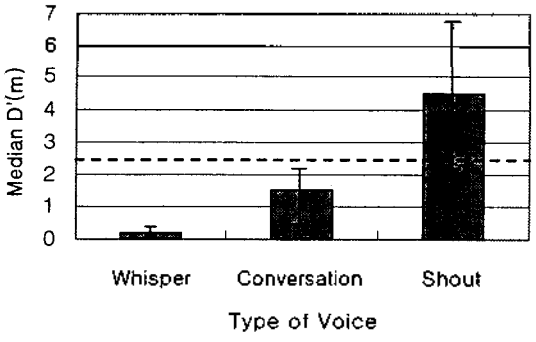

Figure 3. Median perceived distances (D') for the later presentation, using three different styles of voice. Error bars represent ±1 semi-interquartile range. The horizontal dashed line indicates the actual physical distance of the loudspeaker.

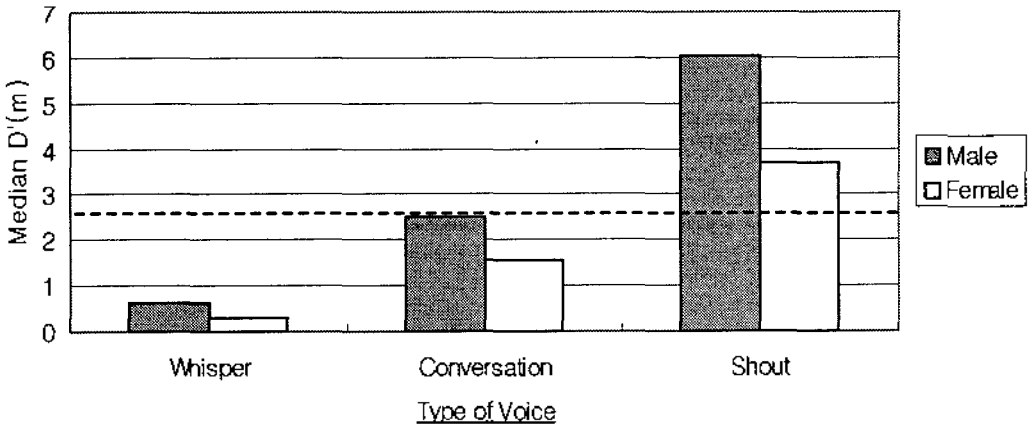

Figure 4. Gender differences in perceived distance (D') to three different styles of voice. The horizontal dashed line indicates the actual physical distance of the loudspeaker.

The findings of a difference based on the sex of the listener (see Figure 4) was expected from a variety of studies from this laboratory which have shown a tendency on the part of male to report slightly larger values, especially under conditions of limited information. This difference may represent a difference in the response values used by the listeners.

The difference associated with the sex of the voice may have two bases. That is, this difference in reported distance is related to the decreased high-frequency content in the deeper male voice, relative to the female voice. However, this does not seem very plausible. Listeners were intentionally denied the opportunity to compare male and female voices and the male-female difference of interest only shows itself for the shouted voices. However, it is important to note that, although small differences exist among the several conditions of the experiment, finding from each sub-grouping reflect the same general result: whispers are perceived as nearer than conversational speech and shouts are perceived as farther.

Because of the interest in finding a stimulus which will be consistently over-perceived with respect to the physical distance of the loudspeaker, some additional analyses were performed on the results for the shouted voices. Since there was a main effect of the listener's sex, two such analyses will be reported. Of 96 reports from women on a shouted voice, 69 women reported that perceived distance was greater than 2.5m (the physical distance of the loudspeaker); 27 reported that perceived distance was less than 2.5m. A sign-test on this difference was significant ($p < .001$). A similar analysis for the 11 reported a distance less than 2.5m. This difference was also significant ($p < .001$). Thus, a strong majority of both male and female listeners reported that the shout was perceived to be farther than the distance of the actual loudspeaker.

V. Discussion and Conclusions

Two points deserve consideration at this point. First, it must be considered the question of whether the differential reports of distance obtained in experiment reflect actual differences in perceived distance or are the result of some cognitive adjustment in the responses based on a generally unchanging perception of distance but a clear knowledge of the expectations of the experimenter. While no final answer is yet possible on this point, the present data provide stronger evidence that a perceptual difference exists than do most previous results.

A second remaining question is whether or not the obvious physical differences in the spectral content of the speech samples could themselves have produced the differences in the reports of perceived distance. Although this was unlikely to occur for the analyses of the first presentation data, it will be important for the future study to evaluate the effects of stimuli which had similar differences in high-frequency content.

Although the results might be criticized because the study was unsure that the sound level and spectral content of all stimulus were equated, there are several conclusions from the experiment. First, listeners clearly report whispers, conversational speech and shouts at systematically different distances, even on initial presentations to separate groups. This implies the efficacy of the familiarity cue to auditory distance. Second, the results do not appear to result from uncontrolled variations in sound level or from the necessary differences in high-frequency content per -se. Third, the over-perception of the distance to the

shouting voice indicates that characteristics of such stimuli may be useful for designing virtual auditory practice. That is, it has proved impossible to create a systematic and stable over-perception of the distance to a source through the manipulation of the acoustic characteristics of the testing room. Therefore, the results suggests that recorded familiar sounds might be able to create the appropriate sorts of perceptual error.

References

1. P. D. Coleman, "An analysis of cues to auditory depth perception in free space," *Psychological Bulletin*, vol. 60, pp. 302-315, 1963.
2. D. H. Mershon and J. N. Bower, "Absolute and relative cues for the auditory perception of egocentric distance," *Perception*, vol. 8, pp. 311-322, 1979.
3. Y-K. Min, "*An adjacency effect in auditory space perception: The effectiveness of relative cues for exocentric distance*," Unpublished Ph.D. Dissertation, North Carolina State University, 1998.
4. A. D. Little, D. H. Mershon & P. H. Cox, "Spectral content as a cue to perceived distance," *Perception*, vol. 21, pp. 405-416, 1992.
5. D. H. Mershon & L. E. King, "Intensity and reverberation as factors in the auditory perception of egocentric distance," *Perception & Psychophysics*, vol. 18, pp. 409-415, 1975.
6. D. H. Mershon, W. L. Ballenger, A. D. Little, P. L. McMurtry & J. L. Buchanan, "Effects of room reflectance and background noise on perceived auditory perception of egocentric distance," *Perception*, vol. 8, pp. 311-322, 1989.
7. E. T. Levy & R. A. Butler, "Stimulus factors which influence the perceived externalization of sound presented through headphones," *Journal of Auditory Research*, vol. 18, pp. 41-50, 1978.
8. R. A. Butler, E. T. Levy & W. D. Neff, "Apparent distance of sounds recorded in echoic and anechoic chambers," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 6, pp. 745-750, 1980.
9. M. B. Gardner, "Distance estimation of 0° or apparent 0° -oriented speech signals in anechoic space," *Journal of the Acoustical Society of America*, vol. 45, pp. 47-53, 1969.
10. P. McGregor, A. G. Horn & M. A. Todd, "Are familiar sounds ranged more accurately?" *Perceptual & Motor Skills*, vol. 61, p. 1082, 1985.
11. R. Guski, "Auditory localization: Effects of reflecting surfaces," *Perception*, vol. 19, pp. 819-830, 1990.
12. W. J. Conover & R. L. Iman, "Rank transformations as a bridge between parametric and nonparametric statistics," *American Statistician*, vol. 35, pp. 124-133, 1981.
13. D. H. Mershon & V. L. Lembo, "Scalar perceptions of distance in simple binocular configurations," *American Journal of Psychology*, vol. 90, pp. 17-28, 1977.
14. S. Handel, "*Listening. An introduction to the perception of auditory events*," Cambridge, MA: MIT Press, pp. 133-162, 1989.
15. D. Rostolland, "*Acoustic features of shouted voice*," Acustica, vol. 50, pp. 118-125, 1982.
16. I. R. Titze, "Physiologic and acoustic differences between male and female voices," *Journal of the Acoustical Society of America*, vol. 85, pp. 1699-1707, 1989.
17. C. A. Fowler, "Auditory perception is not special: We see the world, we feel the world, we hear the world," *Journal of the Acoustical Society of America*, vol. 89, pp. 2910-2915, 1991.
18. P. Cochran, J. Throop & W. E. Simpson, "Auditory spaciousness: Some further psychacoustical analyses," Journal of the Acoustical Society of America, vol. 80, pp. 533-542, 1986.

▲ Yoon-Ki Min

Yoon-Ki Min received the B.A. degree from Chungnam National University and the M.A. degree from ChungAng University, all in psychology, in 1985 and 1987, respectively. And he obtained Ph.D. degree in psychology from North Carolina State University, Raleigh, USA, in 1998. Since 1999, he has been an instructor at Sejong Unioversity, ChungAng University and Chungnam National University. From 2000, he works as a senior researcher at Korea Research Institute of Jungshin Science. His major research area are ergonomics, auditory perception and multi-modal interface design.