PERCEPTUAL CONSENQUECES OF INCLUDING REVERBERATION IN SPATIAL AUDITORY DISPLAYS

Sasha Devore

Boston University
677 Beacon St.
Boston, MA 02215
sashad@mit.edu

Barbara Shinn-Cunningham

Boston University
677 Beacon St.
Boston, MA 02215
shinn@cns.bu.edu

ABSTRACT

This paper evaluates the perceptual consequences of including reverberation in spatial auditory displays for rapidly-varying signals (obstruent consonants). Preliminary results suggest that the effect of reverberation depends on both syllable position and reverberation characteristics. As many of the non-speech sounds in an auditory display share acoustic features with obstruent consonants, these results are important when designing spatial auditory displays for nonspeech signals as well.

1. INTRODUCTION AND BACKGROUND

Echoes and reverberation (henceforth called "reverberation" throughout this paper) provide a robust cue for sound source distance [1-4]. However, reverberation corrupts the signals reaching the ears of the listener and distorts many acoustic features that normally convey information, including amplitude and frequency modulation. It is therefore important to understand the perceptual consequences of including reverberation when designing spatial auditory displays.

Speech is one example of a complex, dynamic acoustic signal that conveys acoustic information to a listener through abrupt energetic onsets and offsets and frequency transitions. Although speech is highly over-learned (i.e. comprehension is automatic), the acoustic properties that convey information in speech are also used in many non-speech auditory displays. As a result, studies of reverberant speech perception can provide insight into what acoustic cues are easily extracted in the presence of reverberation when a listener is highly trained.

Human recognition and understanding of speech is very robust to signal degradations both because there are many redundant acoustic cues to allow a listener to decode the acoustic signal, and because there are linguistic, contextual limitations on language that provide top-down constraints when parsing an ordinary speech utterance. In order to generalize the results from a speech perception study to those that would arise for arbitrary, non-speech stimuli, it is important to factor-out these speech-specific effects. This is readily achieved by using nonsense speech syllables.

This paper describes how reverberation distorts the acoustic properties of nonsense syllables. Preliminary results from a study of the effects of reverberation and competing noise on nonsense syllable identification are then reported. Results are contrasted with those of a previous study [5] that examined sentence intelligibility in similar listening conditions.

1.1. Acoustic Properties of Consonants

The simplest nonsense speech tokens are syllables containing a vowel and one consonant. The vowel portions of such syllables are harmonic with slowly-changing spectral content arising from changes in the resonances of the vocal tract. In contrast, consonants are characterized by rapid spectral change [6]. This paper is concerned with one particular class of consonants (the obstruents) that contain rapid onsets and offsets important for indentification, which could be detrminentally affected by reverberation.

Reverberation tends to distort a signal by temporally smearing energy and masking a speech segment by reflections of the segment itself as well as of previous segments [7]. As consonant identification is largely based on detection of rapid changes in the spectrum of the signal, masking spectral modulations causes confusions [8, 9]. In addition, the temporal smearing of speech reduces amplitude modulations at different frequencies [10]. Reverberation can degrade consonant perception by smoothing out the envelope modulations that carry information about the abrupt energetic onsets and offsets [11, 12]. The temporal relationship of echoes arriving at an ear affects the amount of modulation reduction and this relationship varies with room characteristics as well as the locations and orientations of both listener and source. The speech perception task described below compares performance in three different acoustic environments in order to explore how perception is affected by different room acoustics.

1.2. Spatial Unmasking of Speech

Spatial unmasking refers to an improvement in detection or identification performance that arises when the target and masking sources are at different spatial locations compared to when the sources are at the same spatial location.

For noise-like maskers, spatial unmasking arises from two main mechanisms: monaural (energetic) effects and binaural processing [13]. For spatially colocated target and

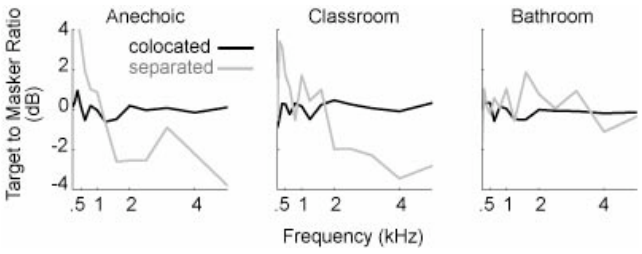

Figure 1. *TMR in 1/3-octave bands needed to achieve broadband TMR of 0 dB at the better ear for colocated and separated sources.*

masker sources, the target-to-masker ratio (TMR) at the two ears will be roughly equal. Displacing either source effectively increases the TMR at one ear (the "better ear") and decreases the TMR at the other ear, leading to improvements in performance due to monaural spatial unmasking. Additionally, even if the levels of the sources are equated to have the same broadband TMR, the TMR typically varies with frequency (see Fig. 1). Because speech information is not uniformly distributed throughout the audible frequency range, these spectral differences can influence speech intelligibility [14]. Even after taking into account the frequency-dependent changes in TMR with changes in target and masker position, additional spatial unmasking can arise when there are differences in the interaural level (ILD) and timing (ITD) differences in the target and masker. The amount of spatial unmasking arising due to such binaural cues depends on the spectral characteristics of the source and target and on the nature of the task [15, 16]].

Reverberation alters ITD and ILD cues as well as spectral shape cues [18] and therefore influences the amount of spatial unmasking in reverberant environments. In small, highly reverberant spaces, the summation of echoes at the two ears decreases ILDs and causes a concomitant decrease in spatial unmasking due to monaural effects. Additionally, as shown in Figure 1, the steady-state TMR is very different in small, highly reverberant environments (right panel) compared to in an anechoic environment (left panel). However, moderate levels of reverberation (center panel) lead to similar steady-state TMRs as in anechoic environments.

2. EXPERIMENT

In order to examine how reverberation influences spatial unmasking of consonants, a study was conducted under headphones. tion through temporal features. The target speech tokens consisted of CV and VC (V= /a/) tokens. Listeners performed a one-interval, nine-alternative, forced-choiced experiment in which they identified which of the nine obstruent consonants /b,d,g,p,t,k,f,v,dh/ was presented, either in quiet or in the presence of a speech-shaped noise masker whose spectrum equaled the average spectra of all target speech tokens. Five normal-hearing subjects were tested on both initial and final consonant identification.

Non-individualized (KEMAR) head-related impulse responses (HRIRs) were used to simulate the target and masker at different spatial locations and in different environments (an ordinary classroom and a bathroom) that varied in the relative levels of reverberant energy reaching the listener (details of the HRIR measurement technique are given in [18]). The classroom HRIRs were also processed (time windowed) to remove all reverberant energy, creating pseudo-anechoic HRIRs. Both target and masker were

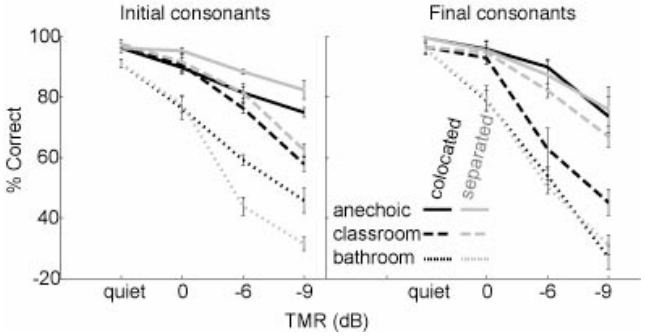

Figure 2. *Monaural performance for initial and final consonants for colocated and separated sources.*

simulated at a distance of 1 m. The target was always simulated from in front of the listener (0°); the masker (when present) was simulated from either 0° or 45° to the right of the subject. Signals were played over a VIA AC'97 Audio Controller soundcard driving Sennheiser HD570 headphones. Performance was measured as a function of broadband root-mean-square TMR at the acoustically "better ear" to estimate the psychometric function. The "better ear" is defined as the ear with the more favorable broadband TMR. In reverberant environments the TMR includes all reverbernant energy as well as direct-sound energy for both target and masker. Subjects were tested both binaurally and monaurally (better ear only) in quiet, with the masker in front of the listener and with the masker at 45˚.

Figure 2 plots percent-correct identification scores for the monaural test conditions as a function of TMR at the better ear (note that chance performance is 1/9 or 11%). In general, for both initial and final consonants, monaural performance decreases both with decreasing TMR and increasing levels of reverberation. However, in quiet (TMR = ∞), performance for initial consonants is statistically better in the classroom than in anechoic space. The consonant confusion matrices reveal that this improvement is due primarily to a reduction in the frequency of /v/ and /dh/ confusions. The moderate level of reverberation in the classroom enhanced subjects' ability to discriminate between these consonants. However, the degradations arising from the interaction of reverberation and noise cause performance to fall more rapidly with decreasing TMR in the classroom than in anechoic space; only in quiet is there a perceptual benefit of classroom reverberation over anechoic listening.

Figure 3 plots the difference between performance for spatially-separated and spatially-coincident target and masker. Because the simulated energy emitted from M was adjusted to fix the overall TMR to the desired value at the better ear, the effects of monaural spatial unmasking are reduced compared to what would happen if the simulated masker was simply displaced in location. Removing the obvious energetic effects that arise from spatially displacing sources, however, reveals spatial unmasking due to spectral tilt in the signals reaching the ears.

Much of the information about consonant identity is conveyed by acoustic cues in the 2 kHz region of the

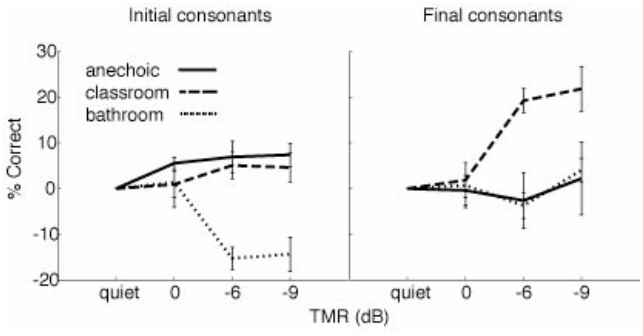

Figure 3. *Effect of spatial separation on monaural performance. Positive values indicate better performance with spatial separation.*

spectrum. For a target and displaced masker, the TMR in one-third octave bands is larger at higher frequencies, even when broadband TMR is fixed, due to head-shadowing of energy above 1.5 kHz in the displaced source signal. The spatial unmasking arising from this spectral tilt in the signals reaching the better ear (Figure 3) depends both on syllable position (initial or final) and room. In the anechoic condition, spectral tilt gives rise to an improvement in performance for initial consonant identification but not for final consonant identification. In the classroom there is consistent spatial unmasking due to spectral tilt for both syllable positions. In the bathroom there is no spatial unmasking for final consonants; however for initial consonants there is actually 'spatial masking:' performance is *worse* when target and masker are spatially separated than when they are at the same location.

Figure 4 compares percent-correct scores for binaural and monaural conditions. Surprisingly, results show no consistent increase in spatial unmasking due to binaural processing. Although with spatially-separated target and masker, binaural performance is generally better than monaural performance, this improvement is roughly the same for monaural conditions and is probably due to the energetic effects discussed above. "Traditional" binaural processing contributions can explain results only when the difference of binaural minus monaural performance is both positive (i.e., binaural performance is superior to monaural performance) and this difference is larger when target and masker are spatially separated than when target and masker are at the same location. In the anechoic condition there appears to be a large contribution from binaural processing to the spatial unmasking of final consonants. However, this effect is primarily due to an improvement for only one of the five subjects. For all other conditions, the binaural advantage is similar when target and masker are spatially separated and when they are at the same location.

Although binaural processing does not contribute to the spatial unmasking of the tested consonants, there is a distinct binaural advantage over monaural listening in both reverberant environments, even when target and masker arise from the same spatial location. In the classroom, the binaural advantage is actually greater when target and masker are in the same location than when they are separated. However, monaural performance is worse for colocated target and masker in this condition; thus, there is more room for improvement with binaural listening. This explanation is consistent with the binaural advantages in the bathroom for initial consonants; binaural performance improves most when the monaural performance is worst. Because monaural

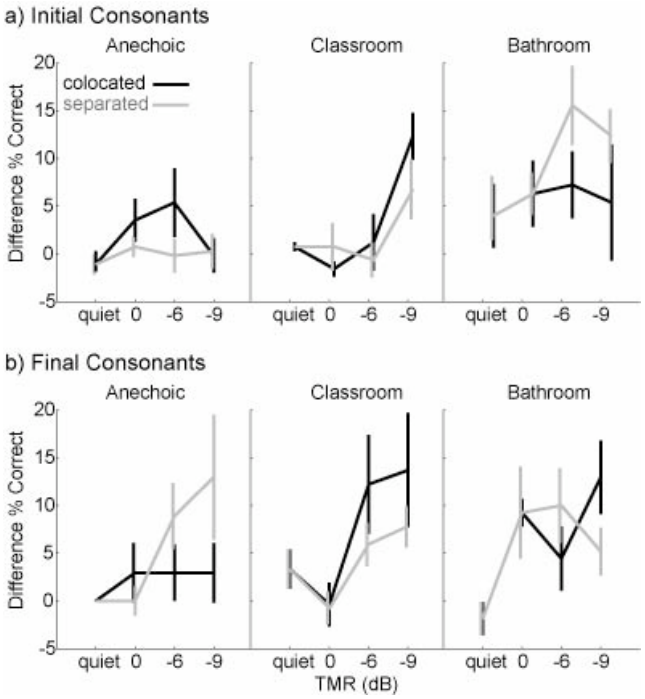

Figure 4. *Binaural advantage (binaural-monaural) for a) initial and b) final consonants.*

performance in the bathroom is independent of spatial configuration for the final consonants, it is also not surprising that the binaural advantage does not depend on spatial configuration in the bathroom. Many discussions of spatial unmasking assume that binaural processing contributions depend on differences at the ears due to spatial location [13]. However, the current results show little evidence of spatial advantage that is mediated by this type of binaural processing.

3. CONCLUSIONS

Results from this study show that improvements in the ability to discriminate among the tested obstruent consonants depend on the acoustics of the listening environment. In a classroom with moderate levels of reverberation, there are clear effects of energetic masking due to changes in the spectral content of the masker reaching the listener's better ear with changes in masker location. However, in the bathroom, where there is more reverberant energy, spatial displacement of the target and masker actually decreases identification performance. In the tested reverberant environments, there is essentially no spatial unmasking beyond monaural effects. However, in these reverberant conditions, binaural performance is generally better than monaural performance. This finding is consistent with previous studies of binaural and monaural speech discrimination of reverberant speech, e.g. [19]. The observed binaural advantage may be due to a statistical decorrelation

of the signals at the two ears (due to asymmetry in the echoes in the room), effectively providing two independent looks at target and masker, one at the left and one at the right ear.

In this study the "beter ear" is defined as that ear with the more favorable steady-state TMR. The independent arrival of echoes at the ears may actually cause temporal fluctuations in the short-time definition of the "better ear." In other words, in the current study, the ear defined as the "better ear" is not necessarily the "better ear" at all instants in time. It is possible that the binaural advantage in reverberant environments arises from a cross-channel integration mechanism in the auditory system integrates information from the "better ear" as it changes from side to side with the arrival of echoes. These kinds of binaural advantages are very different from the binaural advantages normally discussed in the literature, as they do not appear to be due to explicit comparisons between the signals at the two ears [13] or to attending to one particular spatial location [20, 21].

Acoustic properties of initial and final consonants differ due to differences in speech production. Thus, it is not surprising that spatial unmasking and binaural advantages also vary with syllable position. Previous studies in our laboratory demonstrate that binaural processing does contribute to spatial unmasking of nonsense *sentences* in both anechoic and reverberant environments [5]. As discussed above, there are a number of additional acoustic and conextual (e.g. lexical, syntactic) cues available in a sentence perception task. Thus, results from the current study may be more indicative of the effects of reverberation on perception of acoustic signals that are non-linguistic, such as might be used in spatial auditory displays. The current results show that the effect of reverberation on perception of rapid temporal acoustic events depends on the specific environment (i.e. the particular reverberation algorithm used in a spatial auditory display). These results can help guide the design of spatial auditory displays by helping determine what amount of reverberation can be included (to improve realism, provide distance cues, etc.) without perceptually degrading the source signal or destroying important spatial unmasking effects.

4. ACKOWLEDGMENTS

This project was supported by grants from the Air Force Office of Scientific Research and the Alfred P. Sloan Foundation. Steve Colburn and Nat Durlach provided valuable insights in discussions of this work.

5. REFERENCES

[1] D.R. Begault, et al., "Direct comparison of the impact of head-tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source," *J. Aud. Eng. Soc.*, vol. 49, pp. 904-916, 2001.

[2] P. Zahorik, "Loudness constancy with varying sound source distance," *Nature Neuroscience*, vol. 4, pp. 78-83, 2000.

[3] B.G. Shinn-Cunningham, S.G. Santarelli, and N. Kopco, " Tori of confusion: Binaural localization cues for sources within reach of a listener," *J. Acoust. Soc. Amer.*, vol. 107, pp. 1627-1636, 2000.

[4] D.H. Mershon and J.N. Bowers, "Absolute and relative cues for the auditory perception of egocentric distance," *Perception*, vol. 8, pp. 311-322, 1979.

[5] B.G. Shinn-Cunningham, "Speech intelligibility, spatial unmasking, and realism in reverberant spatial auditory displays," presented at ICAD, 2002.

[6] K. Stevens, "Acoustic correlates of some phonetic categories," *J. Acoust. Soc. Amer.*, vol. 68, pp. 836-842, 1980.

[7] A.K. Náblek, T.R. Letowski, and F.M. Tucker, "Reverberant overlap- and self-masking in consonant identification," *J. Acoust. Soc. Amer.*, vol. 86, pp. 1259-1265, 1989.

[8] S.A. Gelfand and S. Silman, "Effects of small room reverberation on the recognition of some consonant features," *J. Acoust. Soc. Amer.*, vol. 66, pp. 22-29, 1979.

[9] K. S. Helfer, "Binaural Cues and Consonant Perception in Reverberation and Noise," *J. Speech Hear. Res.*, vol. 37, pp. 429-438, 1994.

[10] T. Houtgast and H.J.M. Steeneken, "A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria," *J. Acoust. Soc. Amer.*, vol. 77, pp. 1069-1077, 1985.

[11] S. Greenberg and T. Arai, "The relation between speech intelligibility and the complex modulation spectrum," presented at 7th International Conference on Speech Communication and Technology, 2001.

[12] T. Chi. et al., 1999. 106(5): p. 2719-2732., "Spectro-temporal modulation transfer functions and speech intelligiblity," *J. Acoust. Soc. Amer.*, vol. 106, pp. 2719-2732, 1999.

[13] [P. M. Zurek, "Binaural advantages and directional effects in speech intelligibility," in *Acoustical Factors Affecting Hearing Aid Performance*, G. Studebaker and I. Hochberg, Eds. Boston, MA: College-Hill Press, 1993.

[14] H.K. Dunn and S.D. White, "Statistical Measurements on Conversational Speech," *J. Acoust. Soc. Amer.*, vol. 11, pp. 278-288, 1940.

[15] N. I. Durlach and H. S. Colburn, "Binaural phenomena," in *Handbook of Perception*, vol. 4, E. C. Carterette and M. P. Friedman, Eds. New York: Acad. Press, 1978, pp. 365-466.

[16] A.W. Bronkhorst, "The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions," *Acustica*, vol. 86, pp. 117-128, 2000.

[17] B.G. Shinn-Cunningham, "Learning reverberation: Considerations for spatial auditory displays," presented at Proceedings of the ICAD, Atlanta, GA, 2000.

[18] T. J. Brown, *Characterization of acoustic head-related transfer functions for nearby sources*. Unpublished M. Eng. Thesis. Cambridge, MA: Massachusetts Institute of Technology. Electrical Engineering and Computer Science Department., 2001.

[19] S.A. Gelfand and I. Hochberg, "Binaural and Monaural Speech Discrimination under Reverberation," *Audiology*, vol. 15, pp. 72-84, 1976.

[20] R. L. Freyman, U. Balakrishnan, and K. Helfer, "Release from informational masking in speech recognition," presented at MidWinter Meeting of the ARO, St. Petersberg Beach, FL, 2000.

[21] D. S. Brungart, "Information and energetic masking effects in the perception of two simultaneous talkers," *J. Acoust. Soc. Amer.*, vol. 109, pp. 1101-1109, 2001.