Spatial hearing advantages in everyday environments

Barbara G. Shinn-Cunningham
Boston University Hearing Research Center
Departments of Cognitive and Neural Systems and Biomedical Engineering
Boston University, Boston, MA 02215
617-353-5764
shinn@cns.bu.edu

Spatial auditory cues are not a dominant factor in human auditory scene analysis (i.e., in "parsing" the sound reaching the ears to determine the number and spectral content of competing sound sources) [1-5]. However, spatial hearing is very important for *understanding* a target source in an environment that has multiple sound sources [6-9]. Resolving this apparent paradox is critical for understanding how human listeners operate in difficult conditions, for instance when there is a heavy workload and there are competing demands on attention. Such knowledge is very important for designing effective auditory displays and other human-machine interfaces.

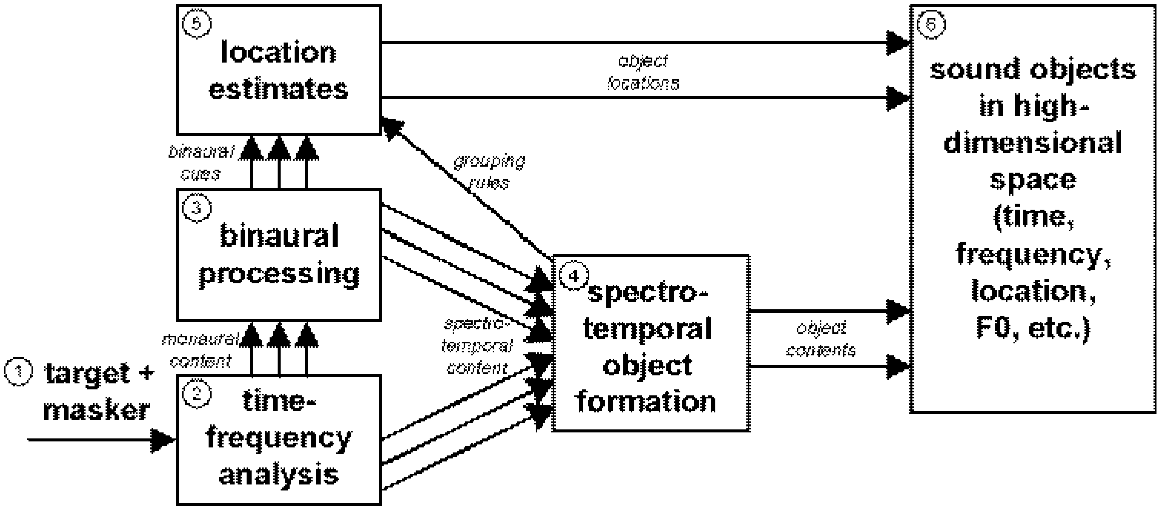
Traditional views of the benefits of spatial hearing [10] fail to explain this contradiction as well as other observations, such as the 1) relatively poor ability of hearing impaired listeners to parse and understand speech in situations with competing sources and/or reverberation [11], 2) relatively large inter-subject differences in performance on tasks involving "informational masking" compared to tasks in which the masker is dissimilar from the target [11-13], and 3) very large improvements in speech intelligibility that can arise when similar competing talkers arise from different locations compared to when they are in the same location [6-9, 14, 15]. This short paper provides a preliminary conceptual framework that unifies these seemingly contradictory results by isolating and identifying multiple ways in which spatial hearing impacts the ability to listen to competing, simultaneous sound sources.

It has long been known that spatial separation of a target from an interfering source (a masker) improves a listener's ability to detect and understand the content of the target (a phenomenon known as "spatial unmasking;" e.g., see [16-18]; recent reviews include [10, 19]). Much of this improvement can be attributed to simple acoustic effects: spatially separating the target and masker generally increases the target-to-masker energy ratio (TMR) at one of the two ears. Because speech intelligibility improves with TMR, the improvement in TMR in one ear leads directly to an improvement in performance. The acoustic TMR varies with frequency because the acoustic interaction of the head and body of the listener with an impinging sound wave varies with the sound wavelength. Thus, the TMR changes more with spatial location of target and masker at high frequencies than at low frequencies. In the most extreme cases (i.e., when one of the sources is very close to the listener), the TMR at moderate frequencies important for speech perception can change by as much as 25 dB (enough so that speech intelligibility could rise from 0% to nearly 100% words correct). While of huge practical importance, changes in acoustic TMR are relatively easy to understand, analyze, and predict using existing models and measurements of the acoustic interaction of sound and a listener's head and body [10, 20].

Especially for a target near threshold, binaural spatial processing (combining information across the two ears) provides an additional important improvement in performance equivalent to increasing the target audibility (effectively increasing the TMR). While the magnitude of this change (on the order of 3-6 dB) is small relative to changes in acoustic TMR, even this change can have a dramatic impact on the ability to understand speech near threshold (e.g., an increase from 10% - 60% words understood) [16-18, 21, 22].

Both changes in acoustic TMR and changes in the effective TMR through binaural processing are effects that have been well characterized by traditional studies of spatial unmasking. For conditions when there is a target talker, a steady-state masking noise, and no room reverberation, current models of spatial hearing are very good at predicting the effects of spatial unmasking on speech intelligibility. However, in many common situations, both the target and a qualitatively similar masker are audible, but difficult to separate perceptually. In these cases, the listener may actually hear some or all of the components in the target, but have difficulty attending to these components in the face of the masker competition. Under these circumstances, perceived differences in the locations of target and masker can improve the ability to attend to and understand the target message [6-9, 11, 23, 24]. In fact, in these situations, spatial separation can cause extraordinarily large changes in performance that even exceed the effects that have been the focus of traditional study (i.e., changes in the acoustic TMR and the effective TMR). Recent work suggests that this form of spatial hearing advantage arises because a listener can selectively attend to a source from a particular location and ignore a competing source from a different direction.

One final factor that has a large impact on spatial hearing advantages in typical environments is the effect of room acoustics [21, 25-29]. Echoes and reverberation alter nearly all acoustics aspects of the signals reaching the ears (spectro-temporal properties, TMR, binaural cues). As a result, the importance of spatial hearing for spatial unmasking differs in “everyday” environments compared to the anechoic conditions under which many psychophysical tests have been performed. Reverberation and echoes degrade spatial unmasking advantages for many traditional test conditions, but are less detrimental on tasks that involve similar-quality, simultaneous talkers and spatial attention. Further, while existing binaural processing models predict spatial unmasking under certain circumstances (e.g., a speech target in the presence of a steady-state, noise masker in anechoic space), they cannot account for the effects of reverberation on speech intelligibility or spatial unmasking.



In order to gain insight into how spatial hearing influences task performance in everyday listening conditions, it is helpful to consider how sound is processed in the auditory system. The spatial auditory pathway is organized in a very hierarchical manner. Even a cursory consideration of its structure suggests that spatial hearing may influence auditory processing at many different stages of processing, and that the importance of spatial hearing for a particular task depends on the nature of the stimuli being presented and the task being performed. The figure above presents our conceptual framework for understanding how spatial hearing influences the ability to listen to a target in the presence of a masking source.

Spatial separation first influences the acoustic TMR through physical interactions, external to the listener (1). Acoustic information is then analyzed neurally to extract spectro-temporal content in the monaural signals reaching the ears (2). This information is processed binaurally in the brainstem (3). We hypothesize that this low-level binaural processing provides information to two parallel processing stages. In one stage, spectro-temporal content of the signals reaching the listener are grouped into acoustic objects by combining spectro-temporal features of the signals reaching the ears (4). We believe binaural processing contributes to this process by revealing spectro-temporal features of a masked signal that may not be audible in a monaural representation (3)-(4). In the grouping stage (4), cues such as harmonicity, common onset, and other features determine how auditory objects are formed. The resulting grouping rules in turn inform the second processing stage (4)-(5) that acts on binaural inputs: the calculation of auditory object location (5). Both location information and spectro-temporal properties of each object are ultimately represented at an even higher level (6). At this stage, selective attention to a particular target attribute (timbre, pitch, loudness, location, level, or other features) can yield large improvements in target understanding by reducing interference from a masking, competing object. Thus, perceived spatial separation can be an extremely helpful and important in mediating competition between simultaneous sources. However, if the masking source is easily segregated from the target through other non-spatial cues (e.g., the masker is steady-state noise that is qualitatively dissimilar to the target in many dimensions), then spatial attention is not needed; other non-spatial cues can be used to focus attention. Spatial separation improves target intelligibility by increasing the TMR (1), increasing the effective TMR (3)-(4), and allowing a listener to attend to an object from a location of interest while ignoring other objects (5)-(6).

The effects of room acoustics on spatial unmasking results can be understood by considering how echoes and reverberation influence the different low- and high-level processing stages in the framework shown in the figure. Echoes and reverberation alter the TMR in the signals reaching the ears (1) and decorrelate the signals at the two ears, reducing the efficiency of binaural processing and the size of the improvement in effective TMR with spatial separation (3). However, echoes and reverberation only modestly degrade spatial perception separation (5). As a result, high-level spatial attention effects that depend on perceived spatial position (6) are not degraded by room acoustics for a normal-hearing listener. The same framework shown in the figure is consistent with other seemingly disconnected observations. For instance, large inter-subject differences generally occur in tasks where high-level spatial attention is critical (5)-(6), whereas inter-subject differences are smaller in tasks that are mediated by lower-level, autonomous processes (3)-(4). This result suggests that the high-level ability to focus attention varies greatly from individual to individual, while lower-level processing is similar across the general population. Hearing-impaired subjects who are faced with a degraded peripheral signal may already suffer from a higher-than-normal workload when trying to understand speech. Thus, conditions in which a listener must focus attention in order to perform a task well (6) (e.g., in reverberant, multi-source environments) may be especially difficult for the hearing-impaired listener, whose resources are already stretched thin.

There is much work remaining to tease apart how spatial hearing influences performance in everyday environments (i.e., with multiple, competing sources that are qualitatively similar, in the presence of reverberation and echoes). Future work should explore the effects of reverberation on peripheral and brainstem representations of monaural and binaural signals; investigate how spectro-temporal features of acoustic signals are grouped and “parsed” by the auditory system; and develop computational models that can account for the effects of reverberation and attention on spatial unmasking in auditory perception. Only with the advent of this type of predictive model can we begin to design effective and efficient auditory displays or determine how to help listeners with hearing impairments cope with everyday listening conditions.

Acknowledgments

Ideas presented in this paper reflect the intellectual input from many collaborators, including TL Arbogast, HS Colburn, NI Durlach, G Kidd, and C Mason.

References

[1] A. S. Bregman, *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge, MA: MIT Press, 1990.

[2] J. F. Culling and Q. Summerfield, "Perceptual separation of concurrent speech sounds: Absence of across-frequency grouping by common interaural delay," *J Acoust Soc Am*, vol. 98, pp. 785-797, 1995.

[3] C. J. Darwin and R. W. Hukin, "Perceptual segregation of a harmonic from a vowel by interaural time difference in conjunction with mistuning and onset asynchrony," *J Acoust Soc Am*, vol. 103, pp. 1080-1084, 1998.

[4] C. J. Darwin and R. W. Hukin, "Auditory objects of attention: The role of interaural time differences," *J Exp Psych: Human Perc Perf*, vol. 25, pp. 617-629, 1999.

[5] C. J. Darwin and R. W. Hukin, "Effectiveness of spatial cues, prosody, and talker characteristics in selective attention," *J Acoust Soc Am*, vol. 107, pp. 970-977, 2000.

[6] T. L. Arbogast, C. R. Mason, and J. Kidd, Gerald, "The effect of spatial separation on informational and energetic masking of speech," *J Acoust Soc Am*, vol. 112, pp. 2086-2098, 2002.

[7] N. I. Durlach, C. R. Mason, B. G. Shinn-Cunningham, T. L. Arbogast, H. S. Colburn, and G. Kidd, "Informational masking: Countering effects of stimulus uncertainty by decreasing target-masker similarity," *J Acoust Soc Am*, in press.

[8] D. S. Brungart and B. D. Simpson, "Within-ear and across-ear interference in a cocktail-party listening task," *J Acoust Soc Am*, vol. 112, pp. 2985-2995, 2002.
[9] D. S. Brungart and B. D. Simpson, "The effects of spatial separation in distance on the informational and energetic masking of a nearby speech signal," *J Acoust Soc Am*, vol. 112, pp. 664-676, 2002.
[10] 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.
[11] T. L. Arbogast, "The effect of spatial separation on informational and energetic masking of speech in normal-hearing and hearing-impaired listeners," unpublished Ph.D. dissertation, *Dept Comm Disorders*, Boston University, 2003.
[12] E. L. Oh and R. A. Lutfi, "Nonmonontonicity of informational masking," *J Acoust Soc Am*, vol. 104, pp. 3489-3499, 1998.
[13] D. L. Neff and T. M. Dethlefs, "Individual differences in simultaneous masking with random-frequency, multicomponent maskers," *J Acoust Soc Am*, vol. 98, pp. 125-134, 1995.
[14] G. Kidd, C. R. Mason, and T. Rohtla, "Binaural advantage for sound pattern identification," *J Acoust Soc Am*, vol. 98, pp. 1977-1986, 1995.
[15] J. Kidd, Gerald, C. R. Mason, T. L. Rohtla, and P. S. Deliwala, "Release from masking due to spatial separation of sources in the identification of nonspeech auditory patterns," *J Acoust Soc Am*, vol. 104, pp. 422-431, 1998.
[16] A. W. Bronkhorst, "The cocktail party effect: Research and applications," *J Acoust Soc Am*, vol. 105, pp. 1150, 1999.
[17] R. Drullman and A. W. Bronkhorst, "Multichannel speech intelligibility and talker recognition using monaural, binaural, and three-dimensional auditory presentation," *J Acoust Soc Am*, vol. 107, pp. 2224-2235, 2000.
[18] A. W. Bronkhorst and R. Plomp, "The effect of head-induced interaural time and level differences on speech intelligibility in noise," *J Acoust Soc Am*, vol. 83, pp. 1508-1516, 1988.
[19] 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.
[20] B. G. Shinn-Cunningham, J. Schickler, N. Kopco, and R. Y. Litovsky, "Spatial unmasking of nearby speech sources in a simulated anechoic environment," *J Acoust Soc Am*, vol. 110, pp. 1118-1129, 2001.
[21] B. G. Shinn-Cunningham, "Speech intelligibility, spatial unmasking, and realism in reverberant spatial auditory displays," Proc Int Conf Aud Display, pp. 183-186, 2002.
[22] M.L. Hawley, R.Y. Litovsky, and J. Culling, "The "cocktail party problem" with four types of maskers: Speech, time-reversed speech, speech-shaped noise, modulated speech-shaped noise," Proc Assoc Res Otolaryng, p. 161, 2000.
[23] N. I. Durlach, C. R. Mason, G. Kidd, T. L. Arbogast, H. S. Colburn, and B. G. Shinn-Cunningham, "Note on informational masking," *J Acoust Soc Am*, in press.
[24] J. Kidd, Gerald, C. R. Mason, T. L. Arbogast, D. S. Brungart, and B. D. Simpson, "Informational masking caused by contralateral stimulation," *J Acoust Soc Am*, vol. in press, 2003.
[25] B. G. Shinn-Cunningham and N. Kopco, "Effects of reverberation on spatial auditory performance and spatial auditory cues," *J Acoust Soc Am*, vol. 111, pp. 2440, 2002.
[26] B. G. Shinn-Cunningham, S. Constant, and N. Kopco, "Spatial unmasking of speech in simulated anechoic and reverberant rooms," Proc Assoc Res Otolaryng, 2002.
[27] J. F. Culling, H. S. Colburn, and M. Spurchise, "Interaural correlation sensitivity," *J Acoust Soc Am*, vol. 110, pp. 1020-1029, 2001.
[28] C. J. Darwin and R. W. Hukin, "Effects of reverberation on spatial, prosodic, and vocal-tract size cues to selective attention," *J Acoust Soc Am*, vol. 108, pp. 335-342, 2000.
[29] J. F. Culling, Q. Summerfield, and D. H. Marshall, "Effects of simulated reverberation on the use of binaural cues and fundamental-frequency differences for separating concurrent vowels," *Speech Comm*, vol. 14, pp. 71-95, 1994.