REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)* 28-01-2007	**2. REPORT TYPE** Final technical report	**3. DATES COVERED** *(From - To)* 1/12/2003-10/31/2006

4. TITLE AND SUBTITLE Auditory and Cross-Modal Spatial Attention	**5a. CONTRACT NUMBER**
	5b. GRANT NUMBER N000140410131
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) Barbara Shinn-Cunningham and Virginia Best	**5d. PROJECT NUMBER**
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Trustees of Boston University Boston, MA 02215	**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Office of Naval Research	**10. SPONSOR/MONITOR'S ACRONYM(S)** ONR
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This final report summarizes results of experiments and theoretical analysis exploring the role of attention, to spatial location or other object attributes, in understanding auditory and auditory-visual objects in complex settings. Work examined how attention affected the ability to understand one (selective attention) or two (divided attention) spoken messages, as well as to identify complex spectro-temporal patterns. Other experiments explored how ambiguous sound mixtures are interpreted and how perceptual objects are formed in complex settings. Theoretical analysis explored the degree to which different acoustic features may help explain the abilities of listeners in complex settings with multiple, competing sources. We find that spatially and non-spatially directed attention, including attention cued through visual signals, enables listeners to better process and understand sound in a complex setting.

15. SUBJECT TERMS

attention, spatial hearing, auditory object formation,

16. SECURITY CLASSIFICATION OF: unclassified			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** Barbara Shinn-Cunningham
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**		10	**19b. TELEPHONE NUMBER** *(include area code)* 617 353 5764

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

Auditory and Cross-Modal Spatial Attention

OBJECTIVES

The original goals of this basic research project were to perform experiments to understand how attention to auditory and multimodal objects affects performance in demanding settings, especially those in which competition for attention limits human abilities. Aim 1 was to test the effect of spatial separation of sources on selective and divided auditory attention, including whether knowledge of spatial configuration affects performance. Aim 2 was to test whether selective attention to space suppresses responses to objects at unattended locations. Aim 3 was to explore the similarities between attention to spatial location and other features of an auditory object. Aim 4 was to determine whether traditional models of spatial hearing can account for the effect of spatial auditory attention.

COMPARISON BETWEEN ACTUAL ACCOMPLISHMENTS AND GOALS

During the course of the three-year grant period, we made excellent progress on all four of the original aims. However, early in the grant period, we realized that the ability of a listener to attend to an audio or audio-visual object depends critically on how a mixture of sounds is perceptually organized into objects. As a result, we redirected some of our efforts to a new goal (Aim 5), investigating the more basic question of how listeners parse an ambiguous mixture of sound coming from multiple sound sources into perceptual objects to which they can attend. Accomplishments in all five areas are summarized in subsequent sections of this report.

APROACH

Behavioral experiments were conducted to explore the ability of listeners to separate, understand, and identify messages from competing auditory objects (using human talkers, songbird calls, and complex harmonic spectrotemporal patterns). Spatial cues in the acoustic signals were manipulated to explore how spatial information influences performance. In some experiments, multiple loudspeakers were used to present sounds from different locations. In other experiments, realistic spatial cues were simulated using virtual auditory space techniques. In auditory mixtures, sounds add before reaching the listeners ears; thus, competing sounds interfere with one another at the most peripheral representation of sound in the auditory neural pathway. To control for such peripheral effects in some experiments, competing speech signals were processed to reduce spectral overlap in order to focus on the effects of central attentional limitations. In selective attention tasks, listeners reported the identity or content of one source in the mixture. In divided experiments, listeners reported the content of both competing sources. In segregation tasks, perceptual organization of the sound mixture was measured indirectly, by measuring the contributions of ambiguous sound elements to object identity and/or to object location.

ACCOMPLISHMENTS

We investigated how spatial configuration and spatial knowledge affect the ability to hear out and understand an auditory target in a complex environment. Overall, results show that the role of spatial cues in hearing out and understanding auditory sources depends on the complexity of the listening environment. More specifically, spatial configuration of the sources in an environment and a priori knowledge of this configuration play a prominent role in helping the listener, and their importance increases with the complexity of the environment (i.e., increasing with the number and similarity of the sources in the sound mixture; Aims 1 and 2). Results also show that in complex environments, spatial cues play multiple roles in helping listeners operate in complex environments (Aims 1, 2, 3, and 5). Spatial cues help us to separate the acoustic energy in the environment into discrete objects or streams, and attend to an object of interest from a particular location. Comparisons between attention to source location, source timbre, and source level show that any one of these features can be used to select out a source from a mixture, and that different listeners rely on these different cues to different degrees. We showed that spatial cues also influence the formation of auditory objects, particularly when the objects are grouped across time. Moreover, it appears that the way in which an ambiguous sound mixture is parsed into objects depends upon what object is attended. Our analysis demonstrates that past models cannot account for the improvements that we observe in complex settings, when attention (rather than audibility) is the main limitation on performance (Aim 4).

Selective attention

We found that listeners attending to one message in the presence of similar messages benefit from perceived spatial separation between the target source and its competitors. Moreover, once the relative levels of the target and masker(s) at the ears are taken into account, the type of spatial cues giving rise to differences in perceived location has no effect on this spatial gain. In particular, the gain is identical when the separated sources contain "realistic" cues (containing all of the natural spatial cues that arise in the real world), only interaural phase differences (a strong left-right direction cue), or only differences in interaural level and interaural envelope timing (weak cues for left-right direction). This work, published in Acustica united with Acta Acustica in 2005, has important implications for display of auditory information to a human operator, as it suggests that in some environments, even extremely simple spatial cues (e.g., simply delaying the sound at one ear relative to the other) can provide all of the possible benefits of doing more complicated, expensive real-time processing to render auditory events at different locations.

Using a new audiovisual paradigm we showed that simple visual cues are effective at guiding attention to auditory objects in complex multi-source environments (in press in Journal of the Association for Research in Otolaryngology). Lights indicating where to listen for a target embedded in similar masking sources from different directions give consistent improvements in performance. The results demonstrate that knowing where to listen is extremely useful for guiding attention in a complex acoustic scene. On the other hand, cues that indicate *when* to listen are only helpful if the target source is extremely difficult to hear out of the mixture (e.g., when the target talker is a familiar bird song that is embedded in other, unfamiliar birdsongs). For target signals that are relatively salient (such as speech targets embedded in reversed speech maskers), cues for when to listen do not aid performance. This demonstrates that visual cues that are temporally correlated with a target can increase listener vigilance and enhance the segregation of a target from a confusing mixture of sources.

Further experiments investigating selective attention to speech demonstrate that listeners can use non-spatial features such as timbre to focus attention on a desired talker. While overall performance does not improve when sources are spatially separated and listeners are attending to timbre (rather than location), the pattern of errors depends on spatial separation. When sources are close together and listeners are attending to timbre, they are more likely to confuse the two talkers, and report a mixture of key words from both talkers. When sources are separated and listeners are attending to timbre, the overall number of correct responses is the same, but listeners rarely report mixtures of the talkers; instead, listeners are more likely to report all keywords from the wrong message. This result shows that spatial separation helps listeners properly separate sources perceptually, even if space is not being used to guide attention. This result gives very strong evidence that in a complex, confusing situation, spatial auditory displays can provide critical information aiding a listener in organizing the acoustic scene, even if they do not know the location of the source that is most important at a given moment.

Divided listening

In a divided listening paradigm, we examined the effect of spatial separation on the ability of listeners to report both messages in a simultaneous pair of speech sources. Results demonstrate that spatial separation improves divided listening, primarily by enhancing intelligibility of the less intense talker. In another, similar study we directly assess the "cost" of dividing attention between sources of equal intensity (published in the Journal of the Acoustical Society of America in 2006). Competing messages were presented either with a small, moderate, or large spatial separation. In separate blocks, we measured the ability of subjects to report the message from one source or from both sources. We assessed the "cost" of dividing attention by directly comparing dual-task performance to the single-task performance. Results show that there is a small increase in the dual-task cost as the spatial separation between the messages increases. This finding is consistent with a "spatial spotlight" model in which listeners are worse on a divided task when the sources are far apart.

Detailed analyses of these two divided listening experiments suggest that listeners cannot listen to more than one source at a time. Instead, in order to report the messages from both of two simultaneous talkers, listeners appear to select one source to attend actively, and then recall a sensory trace of the stimulus from memory in order to report the "lower priority" talker. Moreover, prior work suggests that such a memory trace degrades rapidly with time, so that the lower-priority message will not be able to be recalled with any accuracy unless listeners can recall this memory trace soon after the message ends. In complex listening situations, a listener will be able to extract the meaning of the source that they actively attend during the stimulus presentation. However, they will only be able to report the lower-priority target message if it the message they are actively attending is short – the longer the messages involved, the less able they will be to recall the lower-priority message. In divided listening, the strategy adopted by listeners depends on their expectation. If one of the talkers is expected to be low in intensity compared to the other talker, listeners pay attention to the talker that they expect to be harder to hear. If listeners are presented with two talkers that are equally intense and are instructed which talker to report first, they appear to actively attend to the source that they must report first. In particular, performance is significantly lower for the "lower priority" source that they report second. This result demonstrates that listeners automatically adopt strategies that are near optimal for the expected listening situation. This conclusion is very important, as it suggests that it is very important for listeners to be given all available information about the listening situation. Such knowledge will be used automatically by the listener to determine how to allocate attention optimally, based on the situation.

Sound source segregation

We have conducted a series of experiments exploring the degree to which spatial cues directly influence how listeners interpret a mixture of acoustic energy (distributed over time and frequency) and form auditory objects. We created sound mixtures in which subjects would hear two distinct objects: a repeating tone and, intermingled in time, a harmonic complex. The two objects were constructed such that there was a target tone that logically could belong to either of the two streams (i.e., its frequency matched the repeating tone and was rhythmically consistent with a tone in that repeating pattern; however, it also was harmonically related to the complex and was turned on and off simultaneously with the complex). We then investigated how changing the spatial properties of the elements influenced the perceived streams. We found that spatial cues had a very large influence on the degree to which the target fell within the repeating tone sequence, but only a modest influence on the degree to which the target was heard as part of the harmonic complex. Even more importantly, we found that these independent measures had little to do with one another: that is, the degree to which the target was heard as part of the tone sequence had little power in predicting the degree to which it was heard in the harmonic complex. These results are consistent with recent views of segregation and streaming, which implicate attention in the process. We believe that the way in which the acoustic mixture is parsed depends on which object the listener attends. Thus, we find that spatial cues have a very strong role in how objects are formed across time, but a weak role in how objects are grouped across frequency. This result has important implications for how spatial cues may be used to reduce interference between competing sound messages: spatial cues can be used to attend to a stream of information over time, but cannot, in isolation, help a listener hear out one sound element from a mixture of sound that is otherwise heard as a single object.

We have also measured how ambiguous sound elements contribute to the perceived locations of competing objects. These results show a dissociation between how sound elements contribute to perceived object content or identity, versus how they contribute to perceived object location. In particular, we find that sound elements may not be perceived as part of a particular object, yet can still contribute to that object's perceived location.

IMPACT/APPLICATIONS

In most modern command and control environments, highly trained human operators are expected to deal with extremely complex scenarios and process large amounts of information. One of the limiting factors in such settings is the human's ability to simultaneously monitor, prioritize, and react to competing sources of information. By investigating how stimulus attributes such as spatial cues influence the ability of the human to sort out and separate competing sound sources and focus attention on a source of interest, new insights into the perceptual and cognitive capabilities of the human operator will be gained. Such knowledge is critical for designing displays that allow a human operator to cope with multiple, competing sources of information in natural, effective ways.

PUBLICATIONS

Journal articles

Best V, Ozmeral EJ, Gallun FJ, Sen K, Shinn-Cunningham BG. Spatial unmasking of birdsong in human listeners: Energetic and informational factors. J Acoust Soc Am 2004; 118(6):3766-73.

Shinn-Cunningham BG, Ihlefeld A, Satyavarta, and Larson E. Bottom-up and top-down influences on spatial unmasking. Acta Acust united Acustica 2005; 91:967-9.

Durlach NI, Mason CR, Gallun FJ, Shinn-Cunningham BG, Colburn HS, and Kidd G Jr. Informational masking for simultaneous nonspeech stimuli: Psychometric functions for fixed and randomly mixed maskers. J Acoust Soc Am 2005; 118:2482-97.

Colburn HS, Shinn-Cunningham BG, Kidd G, Durlach N. The perceptual consequences of binaural hearing. Int J Audiology 2006; 45:S34-44.

Best V, Gallun FJ, Ihlefeld A, Shinn-Cunningham BG. The influence of spatial separation on divided listening. J Acoust Soc Am 2006; 120:1506-16.

Best V, Ozmeral EJ, and Shinn-Cunningham BG. Visually-guided attention enhances target identification in a complex auditory scene. J Assoc Res Otolaryng, in press.

Conference papers

Shinn-Cunningham BG and Ihlefeld A. Selective and divided attention:

Extracting information from simultaneous sound sources. Proceedings of the International Conference on Auditory Display, Sydney, Australia, 6-9 July 2004.

Shinn-Cunningham BG (2004). The perceptual consequences of creating a realistic, reverberant 3-D audio display. International Congress on Acoustics, Kyoto, Japan, 4-9 April 2004.

Best V, Ihlefeld A, and Shinn-Cunningham BG. Effect of spatial layout in a divided attention task. Proc. International Conference on Auditory Display, Limerick, Ireland, 6-9 July 2005.

Shinn-Cunningham BG. Influences of spatial cues on grouping and understanding sound. Proceedings of the Forum Acusticum, Budapest, Hungary, 29 August – 2 September 2005.

Shinn-Cunningham BG, Best V, Dent ML, Gallun FJ, McClaine EM, Narayan R, Ozmeral EJ, and Sen K. Behavioral and neural identification of birdsong under several masking conditions. Proc International Symposium on Hearing, Cloppenberg, Germany, 18 – 23 August 2006.

Conference abstracts

Shinn-Cunningham, BG, A Ihlefeld, Satyavarta, and E Larson (2005). "Spatial auditory attention," First Indo-American Frontiers of Science Symposium sponsored by the National Academies of Science, Bangalore, India, 8-10 January 2005.

Ihlefeld, A, E Larson, and BG Shinn-Cunningham (2005). "Spatial unmasking at a spectrally sparse cocktail party," Mid-Winter Meeting of the Association for Research in Otolaryngology, New Orleans, LA, 19-24 February 2005.

Ozmeral, E, V Best, F Gallun, K Sen, and BG Shinn-Cunningham (2005). "Identifying a bird in a chorus: How target and masker statistics influence spatial unmasking," Mid-Winter Meeting of the Association for Research in Otolaryngology, New Orleans, LA, 19-24 February 2005.

Lee, AKC, B Shinn-Cunningham, and A Oxenham (2005). "The missing target: Evidence of a tone's inability to contribute to the auditory foreground," Mid-Winter Meeting of the Association for Research in Otolaryngology, New Orleans, LA, 19-24 February 2005.

Satyavarta and BG Shinn-Cunningham (2005). "Contribution of binaural mechanisms to spatial unmasking," Computational and Systems Neuroscience meeting, Salt Lake City, UT, 17-20 March 2005.

Shinn-Cunningham, BG (2005). "The problem of separating and localizing sound objects," Workshop on Auditory Processing: Localization and Separation, Snowbird, UT, 22 March 2005 [invited talk].

Best, V, A Ihlefeld and BG Shinn-Cunningham (2005). "The effect of spatial configuration in a divided attention task," Journal of the Acoustical Society of America, 117, 2562.

Narayan, R, V Best, E Ozmeral, BG Shinn-Cunningham and K Sen (2005). "Neural discrimination of complex stimuli in the presence of masking sounds," Annual Meeting of the Society for Neuroscience, Washington, DC, 12-16 November.

Ozmeral, E, V Best, and BG Shinn-Cunningham (2006). "Simple visual cues enhance the identification of target sounds in complex auditory scenes," Annual Meeting of the Australian Neuroscience Society, Sydney, Australia, 31 January – 3 February 2006.

Ozmeral, E, V Best, and BG Shinn-Cunningham (2006). "Enhanced target identification in a complex auditory scene via visual cueing," Mid-Winter Meeting of the Association for Research in Otolaryngology, Baltimore, MD, 5-9 February 2006.

Best, V, P Bengani, FJ Gallun, and BG Shinn-Cunningham (2006). "The cost of dividing auditory attention between two spatial locations," Mid-Winter Meeting of the Association for Research in Otolaryngology, Baltimore, MD, 5-9 February 2006.

Lee, AKC, R Cusack, RP Carlyon, and BG Shinn-Cunningham (2006). "Evidence for an effect of attention on the buildup of across-frequency streaming," Mid-Winter Meeting of the Association for Research in Otolaryngology, Baltimore, MD, 5-9 February 2006.

Ihlefeld, A and BG Shinn-Cunningham (2006). "Listening strategies affect the benefit of spatial separation in informational masking," Mid-Winter Meeting of the Association for Research in Otolaryngology, Baltimore, MD, 5-9 February 2006.

Shinn-Cunningham, BG (2006). "Spatial auditory attention," Journal of the Acoustical Society of America, 119, 3416, [invited].

Ihlefeld, A S Sarwar, and BG Shinn-Cunningham (2006). "Spatial uncertainty reduces the benefit of spatial separation in selective and divided listening," Journal of the Acoustical Society of America, 119, 3417.

Lee, AKC, A Deane-Pratt, D McAlpine, and BG Shinn-Cunningham (2006). "Spatial cues are used differently for localizing and identifying the same attended auditory object" Journal of the Acoustical Society of America, 119, 3417.

Lee, AKC and BG Shinn-Cunningham (2006). "Ambiguous sound elements contribute differently to auditory object identity and object location," British Society of Audiology, Chester, UK, 19-21 September 2006.

Lee, AKC and Shinn-Cunningham (2006). "Measuring the distinctiveness of auditory objects using perceived location", Annual Auditory Perception, Cognition, and Action Meeting, Houston, 16 November.

Presentations

Shinn-Cunningham, BG (2004). "Spatial processing and spatial perception in everyday settings," talk presented at the Workshop on Auditory Processing of Vocalizations and other Complex Sounds, Cold Spring Harbor, New York, 24 March 2004.

Shinn-Cunningham, BG (2004). "Auditory attention: Benefits of spatial separation and costs of spatial uncertainty," presented at the Office of Naval Research Workshop on Attention and Complex Processes, Newport, RI, 6-7 May 2004.

Shinn-Cunningham, BG (2004). "Bottom-up and top-down influences of auditory spatial processing," talk presented at the Summer School on Object Formation in Audition and Vision: Bottom-up and Top-Down Processing, Oldenburg, Germany, 18-22 August 2004.

Shinn-Cunningham, BG (2004). "Spatial attention and source segregation," talk presented at the Workshop on Source Segregation, Hanse Institute, Delmenhorst, Germany, 22-24 August 2004.

Shinn-Cunningham, BG (2004). "Acoustic source separation in the brain: How we use two ears," talk presented at the University of New Hampshire College of Engineering and Physical Sciences Frontiers lecture series, Durham, NH, 6 December 2004.

Shinn-Cunningham, BG (2005). "Understanding sound in the complex, everyday world," talk presented at the Tata Institute for Fundamental Research, Mumbai, India, 12 January 2005.

Shinn-Cunningham, BG (2005). "Understanding sound in the complex, everyday world," talk presented at the Indian Institute of Technology, Mumbai, India, 13 January 2005.

Shinn-Cunningham, BG (2005). "Understanding sound in the complex, everyday world," talk presented at the National Brain Research Center, New Delhi, India, 14 January 2005.

Shinn-Cunningham, BG (2005). "Influences of spatial cues on separating and understanding sound sources in divided and selective attention tasks," presented at the Office of Naval Research Workshop on Attention and Complex Processes, Washington, DC, 10-12 May 2005.

Shinn-Cunningham, BG (2006). "Space is just another auditory feature," presented at the workshop on "Difficult Issues in Auditory Scene Analysis," held in conjunction with the Computational and Systems Neuroscience meeting, to be held in Park City, Utah, 12 March 2006.

Shinn-Cunningham, BG (2006). "Searching for cross-modal analogies: Spatial attention in complex auditory scenes," presented at the Visual Attention Seminar Series, Brigham and Women's Hospital, 15 March 2006.

Shinn-Cunningham, BG (2006). "Divide and conquer: Attention in complex auditory scenes," talk presented in the Seminar Series of the Institute for Hearing, Speech and Language and the Department of Speech-Language Pathology and Audiology, Northeastern University, Boston, 23 March 2006.

Shinn-Cunningham, BG (2006). "Communicating in a natural cocktail party: Relating human and avian behavior to neural response," colloquium for the Center For Adaptive Systems, Department Of Cognitive And Neural Systems and the Center Of Excellence For Learning In Education, Science, And Technology (CELEST), Boston University, Boston, 14 April 2006.

Shinn-Cunningham, BG (2006). "Auditory attention and spatial hearing," presented at the Telluride Neuromorphic Engineering Workshop, Telluride, Colorado, June 2006.

Shinn-Cunningham, BG (2006). "Spatial hearing considerations in designing effective hearing aids," presented at the Oticon Research Group, Copenhagen, Denmark, August 2006.

Shinn-Cunningham, BG (2006). "How attention influences what we hear," presented at the Cognition, Brain, and Behavior Research Seminar, Cambridge, 5 October 2006.