BOSTON UNIVERSITY

GRADUATE SCHOOL OF ARTS AND SCIENCES

Dissertation

SPATIAL HEARING, AUDITORY SENSITIVITY, AND PATTERN RECOGNITION IN NOISY ENVIRONMENTS

by

NORBERT KOPČO

Ing., Technická Univerzita Košice, Slovakia, 1996

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2003

Approved by

First Reader

Barbara G. Shinn-Cunningham, Ph.D.
Assistant Professor of Cognitive and Neural Systems and
Biomedical Engineering

Second Reader

Gail A. Carpenter, Ph.D.
Professor of Cognitive and Neural Systems and Mathematics

Third Reader

H. Steven Colburn, Ph.D.
Professor of Biomedical Engineering

Acknowledgments

Many people contributed to my education and helped me shape my research interests. Barbara Shinn-Cunningham, my primary advisor, introduced me to the field of psychoacoustics. Barb was a great mentor, again and again surprising me by her ability to combine her devotion to science and her family, while being always available for a friendly chat. Gail Carpenter, who advised me in a part of this thesis research, showed me that science requires a lot of persistence and questioning of things that appear to be obvious. Steve Colburn, my third reader, with his broad knowledge of binaural hearing and openness to discuss it at any time would be very influential for me and for this thesis even if he was not a member of my dissertation committee.

I would like to thank all the people in the binaural group in the Boston University Hearing Research Center and in the memory group in the Cognitive and Neural Systems department for creating an exciting and friendly atmosphere. In particular Tim Streeter helped me with research and other obstacles on the long path towards a Ph.D.

There were many people in my home Slovakia who supported me during my studies in Boston. Peter Sinčák, my diploma thesis advisor, had a great influence on my decision to come to Boston, while the students in his research group at Technická Univerzita Košice were a vital link, making sure that I never really left Košice.

I would like to thank my parents, Andrej Kopčo and Ružena Kopčová, for their patience and support throughout the many years of my studies: Vďaka za všetko!

Finally, I want to thank Antje for many things, for example, for making sure that I don't forget that there are other things around besides this thesis.

SPATIAL HEARING, AUDITORY SENSITIVITY, AND PATTERN

RECOGNITION IN NOISY ENVIRONMENTS

(Order No.)
NORBERT KOPČO
Boston University Graduate School of Arts and Sciences, 2003
Major Professor: Barbara Shinn-Cunningham, Assistant Professor of Cognitive and Neural Systems and Biomedical Engineering

ABSTRACT

This dissertation investigates how human perceivers and learning algorithms cope with noisy, complex environments. The first part of the dissertation presents results of psychoacoustic experiments that give insight into human auditory perception of sounds originating near the listener. The second part develops a memory-based computational learning architecture that is resistant to noise while limiting its own internal complexity.

The first part of the dissertation describes results of two psychoacoustic experiments. The first experiment investigates the ability of human listeners to employ binaural auditory processing when detecting a sound from various positions around the listener. The experiment is performed in a simulated noisy anechoic environment, which provides more realistic spatial cues than have been tested in previous headphone studies of masked sound detection while allowing the acoustic signals to be fully characterized. The results support the idea that performance is determined by brainstem auditory processing and is not influenced by any higher-level sound-location-based processing. The second psychoacoustic experiment studies how the listener's position in a room and the listener's experience in a room influence the ability to accurately localize nearby sounds in azimuth and distance. The results show that the listener position in the room influences response variability and that as a listener becomes familiar with a room, passive learning occurs, reducing response variability. Finally, the presence of a nearby wall has a small influence on azimuthal response bias, but no noticeable effect on perceived source distance.

In the second part of the dissertation, a memory-based learning system, called PointMap, is proposed. This system implements new methods of pruning based on informative value of the stored memories. It is shown that such a "forgetting" mechanism can solve the problems of category proliferation and noise sensitivity in many learning systems. Variations on the Adaptive Resonance Theory systems are also investigated.

Both portions of this dissertation demonstrate that learning enables humans and machines to perform robust pattern recognition tasks in the face of noise and uncertainty in the inputs.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ART	Adaptive Resonance Theory
CNN	Condensed Nearest Neighbor
ENN	Edited Nearest Neighbor
HRTF	Head-Related Transfer Function
ILD	Interaural Level Difference
IPD	Interaural Phase Difference
ITD	Interaural Time Difference
MLS	Maximum-Length Sequence
NN	Nearest Neighbor

SPATIAL HEARING, AUDITORY SENSITIVITY, AND PATTERN RECOGNITION IN NOISY ENVIRONMENTS

Chapter 1 Introduction

Every organism and almost every artificial system interacts with the environment that surrounds it. This interaction is based on sensory information received from the environment. Many factors influence the ability to correctly process the input sensory information, including corruption by external noise from various sources, such as low-quality sensors, environmental complexity, or the presence of multiple, simultaneous sources of sensory information. Therefore, separating useful information from noise is a general problem with which all organisms and systems must cope. This dissertation investigates the problem of separating information from noise in two distinct domains. First, human auditory perception is studied in a series of psychoacoustic experiments that investigate how humans detect sounds masked by noise and how they determine the location of sound sources in reverberant rooms. Second, general computational learning algorithms are developed that are able to identify and suppress noise in the inputs received from the environment.

1.1 Spatial hearing

The auditory system in humans and animals processes acoustic signals received at the ears to extract information encoded in the signals (Moore, 1997). The information extracted can include: a message encoded in the signal (linguistic content of speech, emotional content of a melody), the identity of the signal source (the human speaker, a mosquito), or the spatial location of the sound source. Experiments conducted as part of the current thesis examine the mechanisms underlying human spatial hearing (Blauert, 1997). The general goal is to further our understanding of how humans determine spatial location of sound sources, how they use this information when performing various auditory tasks, and how these processes are influenced by the acoustic environment. Spatial hearing allows listeners mainly to perform two kinds of tasks: to localize sources of sounds and to improve perception of sounds masked by other spatially-separated sounds. Performance in both of these kinds of tasks is considered in this dissertation.

Over the last century, spatial hearing has been extensively studied (Gilkey and Anderson, 1997). However, most studies examined perception of sounds coming from sources relatively far from the listener (not within the reach of his/her hands) in anechoic space (Brungart and Durlach, 1999). Moreover, most studies fixed the distance of the sound source and examined sensitivity to source azimuth or elevation without considering source distance (Middlebrooks and Green, 1991). Of course, for sources more than about one meter from the listener in anechoic space, most spatial auditory localization cues do not change with source distance. However, this is not the case for nearby sources. The present studies examine spatial perception of sound sources originating within reach of the listener as a function of source distance.

This dissertation presents results of two spatial hearing studies, introduced separately in the following two sections.

1.1.1 Detection of pure-tone sources masked by noise

One well-known spatial hearing phenomenon is the "cocktail party effect" (Bronkhorst, 2000), which refers to the ability to selectively listen to one sound source and ignore other simultaneous sound sources, particularly when the competing sources are in different spatial positions relative to the listener. This effect has been studied for a variety of complex stimuli (speech, tone complexes, noise bursts). However, no previous study has examined how source distance influences a pure tone target signal in the presence of a broadband noise masker or considered the effect of source distance. On the other hand, there is extensive body of literature and several models that characterize detectability of pure tones in noise when the signals are presented via headphones (Durlach and Colburn, 1978). This literature shows that the binaural cues, i.e., the differences in timing and intensity with which the target tone and the masking noise are presented at the two ears can provide the most salient detection cues for tones in noise when the binaural cues in the tone and noise differ.

The first experiment presented in this thesis measures how detectability of a pure-tone target masked by a noise masker is influenced by the location of the sources when the sources are within reach of the listener. Detection thresholds are measured for various combinations of azimuth and distance of the sources relative to the listener in a simulated anechoic environment. An existing physiologically based model of binaural auditory processing is then combined with acoustic analysis to predict the experimental results.

1.1.2 Localization in reverberant rooms

In real rooms, the acoustic signals reaching the listener are influenced by the sound reflections off walls and other surfaces (Hartmann, 1997). This "reverberation" can degrade directional auditory perception, but it can provide a cue for distance perception (Santarelli, 2000). The acoustic effect of reverberation depends on the position of the listener in the room as well as the position of the source relative to the listener (Brown, 2001). When the listener is in the center of the room, the effect of reverberation is roughly constant independent of the sound source position, because all the walls are relatively far from the listener. When the listener is close to a wall, the effect of reverberation depends on the sound source position due to changes in the relative timing of and direction of incidence of early reflections off the wall. Finally, a recent study (Shinn-Cunningham, 2000) suggests that listeners' performance improves over time when listeners perform localization tasks in such rooms.

The second experiment in this thesis measures localization of nearby sound sources varying in azimuth and distance. In order to investigate the effects of changes in room acoustics on localization performance, each listener is tested in different locations in the room. In order to tease apart the influence of experience on localization, subjects were assigned to one of two groups that differed in the order in which the different room locations were tested. Behavioral results are compared to acoustic measurements of the signals reaching the ears of a KEMAR acoustic manikin that show how localization cues are affected by room reverberation in different listener locations in the room.

Together these two experiments provide a basis for future research investigating how humans use spatial hearing in everyday listening situations. The results might be also useful in various practical applications, including hearing aids and virtual auditory displays.

1.2 Memory-based learning and Adaptive Resonance Theory

A general pattern recognition task is to design an artificial system that can learn to classify objects into distinct categories based on certain features (e.g., to classify humans as children vs. adults based on height) (Duda, Hart and Stork, 2001). There are many possible strategies a system can use to learn this task. The memory-based learning systems' approach is to store all (or a subset of) the presented exemplars during training. After training, these systems classify a new exemplar into the class of the most similar pattern(s) among the ones stored (Dasarathy, 1991). Memory-based learning systems have been shown to be very accurate, but they have several weaknesses, including large memory requirements and sensitivity to noise. The present thesis proposes a method of pruning based on the information value of each of the stored exemplars to cope with these problems. The information value of an exemplar is defined as a combination of its predictive accuracy and criticality. Pruning allows the system to limit its size on-line during learning, as well as after the training is over. Depending on the choice of a system parameter, the information-value computation and pruning can bias the system to focus on fine detail in the training data or to generalize and eliminate most of the noisy exemplars. The information value computation and pruning methods are implemented in a new incremental learning system called PointMap and evaluated on several benchmark problems.

ARTMAP neural network systems (Carpenter, Grossberg and Reynolds, 1991; Carpenter, Grossberg, Markuzon, Reynolds and Rosen, 1992) share certain characteristics with memory-based learning systems. Instead of retaining individual training exemplars, ARTMAP systems generate a set of hyper-rectangles each of which encodes multiple exemplars from the training set. The similarity between an unknown exemplar and the stored code is evaluated by a match function which is typically insensitive to the position of the exemplar if the exemplar lies within the hyper-rectangle. That is, an exemplar that is within the hyper-rectangle but close to its border and an exemplar in the center are encoded by the hyper-rectangle equally. This may cause sensitivity to noise in the ART systems because noise can cause an exemplar that should be outside the hyper-rectangle to be shifted into it, whereas an exemplar that is in the center of the hyper-rectangle is less susceptible to this kind of noise. The final project of this thesis develops graded signal functions for ARTMAP neural networks. These functions generalize the original signal function so that it is sensitive to a position of the exemplar within the hyper-rectangle.

Both the PointMap pruning methods and the ARTMAP graded signal functions are developed here for specific algorithms, but they can be implemented into most memory-based or ARTMAP systems.

1.3 Organization of dissertation

This dissertation is divided into three main parts. The first part (Chapter 2) contains a review of background information for each of the studies. The second part

describes the two auditory experiments: Chapter 3 describes spatial unmasking of pure tone signals while Chapter 4 and Chapter 5 describe auditory localization in real rooms. The third part describes two pattern recognition studies: Chapter 6 develops the PointMap incremental memory-based learning system and Chapter 7 describes the new signal function for the ART neural networks. Chapters 3-7 were originally written for journal publication, so each includes a review of the literature for that particular study.

Chapter 2 Background

This chapter contains general background for the topics studied in the dissertation. Background material specific to a particular chapter is provided at the beginning of that chapter. Topics covered here include: monaural and binaural cues for detection of sounds in noise and for auditory localization, effects of room reverberation on binaural sound localization cues, and memory-based learning algorithms and Adaptive Resonance Theory (ART).

2.1 Spatial hearing

When a sound is produced, it propagates from its source through the environment until it reaches the listener's ears. The sound received at the ears differs from the sound produced by the source because it is modified by interactions with the listener's body, head, and pinnae (Brungart and Rabinowitz, 1999; Shinn-Cunningham, Santarelli and Kopčo, 2000). In addition, if there are acoustically-reflective objects (for example walls) in the environment, acoustic reflections off these objects are received by the ears along with the "direct" sound. The basis of spatial hearing is in that the listener's auditory system extracts cues about the location of the sound source from the sounds received at the ears and the listener uses these cues to perform various tasks (localization, detection of sounds).

2.1.1 Head-related transfer functions

The transformation of a sound from the source to the ear is constant for a fixed sound source and listener position. The sound source, the environment in which the sound propagates (including the listener, and all objects and walls in the environment), and the ear create a linear system that transforms the input signal (sound produced by the source) into an output signal (the sound received at the ear). This system can be mathematically characterized by its impulse response called the Head-Related Transfer Function (HRTF). The HRTF describes the signal reaching the ear when a broadband impulse is played from a specific source location. This impulse response is sufficient to predict how any sound coming from a specific location is altered as it travels to and impinges on the ear. Because the sound has to travel through a different path to each of the two ears, a pair of HRTFs (the left-ear and the right-ear HRTF) provides complete information about how any sound is received at the ears when originating from a given location.

There are two main applications of HRTFs in hearing research. First, HRTFs can be used to generate virtual auditory environment. That is, by convolving a sound with the HRTF one can simulate how the sound would be received at the listener's ears if it was presented from any location around the listener in any environment. Second, HRTFs can be analyzed to determine what spatial auditory cues are available to the listener when a sound is presented from a specific location in a specific room.

2.1.2 Basic cues for spatial hearing

The human auditory system extracts from the sounds received at the ears two kinds of acoustic cues (Blauert, 1997). "Monaural" cues depend only on the sound

received at each ear separately. ”Binaural” cues depend on comparing the signals received at the two ears. The most important monaural localization cue is the change in the magnitude spectrum of the sound caused by the interaction of the sound with the head, body, and pinna before entering the ear. The most salient binaural cues are differences in the time of arrival (the interaural time difference, ITD, which can be converted into interaural phase difference, IPD) and differences in the intensity (the interaural level difference or ILD).

Monaural cues are more ambiguous spatial cues than binaural cues because the auditory system must estimate what spectral features are due to the original spectrum of the source and what are due to the filtering of the head and body (and thus depend on source location). Although in theory it is not possible to separate effects of the source spectrum from spectral filtering in the HRTF, listeners are familiar with many everyday sounds. Similarly, if an unfamiliar sound is produced several times from various locations, the listener may be able to learn what spectral features are due to source location and what are due to source content.

In contrast with monaural cues, binaural cues are essentially independent of the acoustic characteristics of the original sound. The only prerequisite for availability of these cues is that the source has sufficient spectral energy in the frequency region for which the cue is to be extracted.

2.1.3 Effects of reverberation on spatial cues

When the listener is in a reverberant environment (e.g., in a room) the direct sound received at the ears is combined with multiple copies of the sound reflected off the walls before arriving at the ears. This reverberation acts like noise that deteriorates the spatial cues extracted by the auditory system. On the other hand, reverberation itself can be a spatial cue.

2.1.4 Sound localization

Positions of objects in 3-dimensional space are usually described using either Cartesian (x, y, z) or spherical (azimuth, elevation, distance) coordinates. For studies of spatial hearing, the most natural coordinate system uses a bipolar spherical coordinates (similar to the coordinate system used to describe a position on the globe) with the two poles at the two ears and the origin is the middle point between the ears (Duda, 1997). In this coordinate system the azimuth of an object is defined by the angle between the source and the interaural axis (equivalent to latitude on a globe) and the elevation is defined as the angle around the interaural axis (equivalent to the longitude on a globe). The third spatial dimension is then the distance from the center of the head to the object. Using this coordinate system is natural when discussing spatial hearing because different auditory localization cues map onto these coordinate dimensions in a natural, monotonic manner. The three spatial dimensions are covered by the auditory localization cues as follows (Wightman and Kistler, 1997).

2.1.4.1 Perception of source azimuth

For relatively distant sources, binaural cues are the primary cues for perception in the azimuthal dimension. Specifically, for low-frequency stimuli (below 1-2 kHz), the ITD changes relatively rapidly with source azimuth and the ILD changes relatively slowly with the source azimuth. Unsurprisingly, the perceived azimuth of low-frequency sounds is dominated by the ITD. For high-frequency stimuli (above 1-2 kHz) the ILD changes rapidly with azimuth due to head-shadow effects. Moreover, the ability of the auditory nerve fibers to encode temporal information is lost because the neurons' maximum firing rates are not high enough to enable them to fire in phase with the stimulus. The auditory system weights the ILD more highly when determining the azimuth of high-frequency, distant sources (Strutt, 1907).

This simple dichotomy (ITD for low frequencies, ILD for high frequencies) has limited application for nearby sources as studied here (Shinn-Cunningham et al., 2000). The main difference for nearby sources is that a reliable ILD cue is available even at low frequencies. The ILD cue for distant sources comes mainly from the head shadow effect, that is, from the fact that the head creates an obstruction for the sound traveling to the more distant ear. For nearby sources, the size of the head is comparable to the source distance and the ILD arises primarily due to the difference in the distances from the source to each of the ears.

In theory, the azimuth of a sound source can be determined also monaurally, because the high-frequency components of the sound are attenuated more compared to low-frequency components as the sound source moves contralaterally (away from the ear in the azimuthal dimension). However, compared to the binaural cues this cue is weak and ambiguous.

2.1.4.2 Perception of source elevation

If the human head were a perfect sphere with the ears exactly opposing each other, there would be no binaural cues for elevation because the binaural cues are constant, creating the well-known cones of confusion (Santarelli, Kopčo, Shinn-Cunningham and Brungart, 1999b). However, small head asymmetries may provide a weak binaural elevation cue. The main cue the auditory system uses to determine the elevation of a sound source is the monaural spectrum determined by the interaction of the sound with the pinnae (Wightman and Kistler, 1997). Specifically, there is a spectral notch that moves in frequency from approximately 5 kHz to 10 kHz as the source moves from 0° (directly ahead of listener) to 90° (above the listener's head) and that is thought by some to be the most prominent elevation cue (Musicant and Butler, 1985).

For nearby sources, the asymmetries of the head and of the position of the ears are emphasized more than for distant sources. Therefore, the changes in binaural cues are larger. Still, the monaural cues dominate elevation perception.

2.1.4.3 Perception of source distance

Little is known about exactly how the auditory system determines the distance of a sound source (Santarelli, 2000). For very distant sources in an anechoic environment

the only available distance cues are the changes in the frequency spectrum (Coleman, 1968) or overall level, which can only be used if the listener has a priori knowledge of the sound source. For nearby, lateral sources the ILD changes with source distance and provides a distance cue (Brungart, 1998). Thus, the cones of confusion for distant sources transform for nearby sources into "doughnuts" of confusion (Shinn-Cunningham et al., 2000). In reverberant rooms, the auditory system uses some aspect of reverberation to determine the source distance (Bronkhorst and Houtgast, 1999).

2.1.4.4 Localization in reverberant rooms

Most earlier studies of sound localization were performed in an anechoic chamber and measured performance in only two dimensions: azimuth and elevation (Wightman and Kistler, 1989; Makous and Middlebrooks, 1990; Wenzel, Arruda, Kistler and Wightman, 1993). There are also several studies of localization in reverberant environments (Hartmann, 1983; Rakerd and Hartmann, 1985, 1986; Wagenaars, 1990). In addition, several recent studies measured also the perceived source distance (Bronkhorst and Houtgast, 1999; Santarelli, 2000; Zahorik, 2000). These studies show that in reverberant space, distance perception is more accurate. However, reverberation also causes small degradations in directional localization accuracy (Santarelli, 2000), although performance improves with practice (Shinn-Cunningham, 2000).

In a recent study, Brown (2001) analyzed the effects of reverberation on acoustic characteristics of the perceived sounds. This study showed that reverberation alters the monaural spectrum of the sound as well as the interaural level and phase differences of the signals reaching the listener. These effects depend on the source position relative to the listener as well as on the listener position in the room.

The present study compares acoustic analyses of the signals reaching a listener in a reverberant room to behavioral results from a localization study performed in the room from which the acoustic measurements were taken.

2.1.5 Detection of sounds masked by noise

2.1.5.1 Factors determining detectability of masked targets

When listening for a target auditory signal in the presence of another simultaneous signal (a masker), a listeners' ability to perceive the target is influenced by the target and masker locations. In general it is easier to detect or recognize the target when it is spatially separated from the masking sound compared to the condition when the two sources are located at the same position (Ebata et al., 1968; Saberi et al., 1991; Good et al., 1997; Kidd, Mason, Rohtla and Deliwala, 1998). Three factors contribute to this *spatial unmasking* effect. First, the acoustic signal-to-noise ratio (SNR) at either ear changes with target and the masker location due to both head shadow effects and distance effects. Spatial separation of target and masker can either increase or decrease the SNR at a given ear, depending on the spatial locations of target and masker.

In addition to simple energetic effects due to changes in SNR at the ears, changes in source location lead to changes in the binaural cues due to that source. The auditory system can detect the presence of the target due to changes in the binaural cues in the

target plus masker stimulus compared to the binaural cues in the masker alone. In general, the target influences IPD cues in the target plus masker most when the IPDs in the target and masker are most different; thus, target detection is easiest when target and masker IPDs differ by π. Similarly, detection of an in-phase target masked by an in-phase noise is easiest if the ILD of the masker is 0 and ILD of the target is ∞ (Durlach and Colburn, 1978).

Finally, informational masking can be influenced by the perceived spatial locations of target and masker. While there is no "standard" definition of informational masking, it is used to refer to influences that cannot be ascribed to simple acoustical parameters of the sounds reaching the two ears (e.g., attentional effects, cross-modality influences, etc.). Informational factors have been shown to play an especially important role for tasks involving high levels of uncertainty, e.g., when complex sounds are masked by complex sounds (Kidd et al., 1998) or when speech is masked by speech (Freyman, Helfer, McCall and Clifton, 1999; Hawley, Litovsky and Colburn, 1999).

2.1.5.2 Previous studies of binaural and spatial unmasking

Spatial unmasking has been studied under headphones and in the free field. Data from headphone experiments address many aspects of auditory processing (Durlach and Colburn, 1978; van de Par and Kohlrausch, 1999) and there are several models that successfully explain observed performance (Colburn and Durlach, 1978; Colburn, 1996). Free-field studies, all of which were done with sources relatively far from the subject's head, generally focused on determining the relative contributions of energetic, binaural, and informational factors to performance. Both headphone and free-field studies analyzed unmasking for pure tone, complex sounds, and speech stimuli.

Results of several studies of free-field masking of pure tones are available (Ebata, Sone and Nimura, 1968; Gatehouse, 1987; Santon, 1987; Doll, Hanna and Russotti, 1992; Doll and Hanna, 1995). These studies used a range of frequencies (200 – 6000 Hz), but restricted source locations to the frontal horizontal plane and at a fixed distance at least 1 m from the center of the head. Unmasking of up to 24 dB was observed. Most of this effect was probably due to the energetic effects, although binaural interactions undoubtedly contributed. For pure-tone targets, the role of informational masking is thought to be negligible; at threshold, pure-tone detection is determined by the subject's ability to detect subtle changes in the masker due to the presence of the target, not by the ability to "hear out" the target as a separate auditory event. In general, even for signal levels above threshold where the target is perceived as a separate object, its perceived location is strongly biased by the location of the masking noise (Santon, 1987). On the other hand, Lutfi (1990) argues on a theoretical basis that 22% of the masking observed in traditional tone-in-noise detection experiments is due to informational masking.

Free-field masking of click-train targets has also been studied (Saberi, Dostal, Sadralodabai, Bull and Perrott, 1991; Good, Gilkey and Ball, 1997), leading to spatial unmasking of up to 20 dB (similar to the results for pure-tone targets). No quantitative modeling of these data has been performed. However, Good et al. (1997) suggest that these data could be predicted by determining the frequency band with most favorable SNR at one of the ears and estimating the binaural cues in that frequency band. Several

studies (Watson, Kelly and Wroton, 1976; Kidd, Mason, Deliwala, Woods and Colburn, 1994) have examined the contribution of informational masking to spatial unmasking of complex sounds. In these studies the masker was a complex of masking tones that was randomly varied from presentation to presentation, introducing a factor of uncertainty that made the task of detecting/recognizing the target much harder. For this kind of task, the spatial unmasking was up to 30 dB. These results show that spatial unmasking is very important in informational masking tasks.

A majority of studies of free-field spatial unmasking looked at changes in intelligibility of masked speech (i.e., the cocktail-party effect, Cherry, 1953). These studies (reviewed in Bronkhorst, 2000) show that the relative binaural contribution to unmasking of speech is generally smaller (e.g., 3 dB) than for pure-tone targets. This result is at least partially explained by the fact that the spectral region important for speech understanding (2-5 kHz) does not overlap with the spectral region for which binaural unmasking effects are large (100-1000 Hz). On the other hand, recent studies of spatial changes in intelligibility of masked speech (Hawley et al., 1999; Freyman, Balakrishnan and Helfer, 2000) suggest that informational factors play a significant role in spatial unmasking, especially when the target speech is masked by an interfering speech signal.

2.1.5.3 Models of binaural unmasking

There are several models that can successfully predict performance in binaural and spatial unmasking tasks (Colburn and Durlach, 1978). The Equalization and Cancellation (EC) model (Durlach, 1972) describes binaural detection as a process in which the signals received at the two ears are first equalized by finding the best time-delay to equate the noise in the left and right ear signals, then subtracting the two signals (canceling the noise). Colburn (1977) proposed a physiologically-plausible model to predict detectability of tones and complex sounds masked by noise. Zurek (1993) extended this (Colburn, 1977a) model to predict spatial unmasking of speech masked by noise. While the Zurek model has been developed to predict spatial unmasking of speech, the Colburn model (or any other model) has not been previously applied to quantitatively predict spatial unmasking of non-speech stimuli.

2.1.5.4 Present study

The present study measures spatial unmasking for pure tone targets masked by broadband noise. The study is performed in a simulated anechoic environment, bridging the previous headphone and free-field spatial unmasking studies. In contrast to previous studies, it explores spatial unmasking for nearby sources, and measures unmasking for source separation in distance, not only in azimuth. Also, the energetic and binaural cues are extracted from HRTFs used in the simulation and predictions of the Colburn (1977) model based on these cues are compared to the behavioral data.

In general, the two studies presented in this dissertation contribute to our understanding of how humans cope with noisy, complex environments.

2.2 Memory-based learning and neural networks for pattern recognition

Pattern recognition (Duda et al., 2001) is a task that humans perform continuously, for example, whenever we look around us, listen to radio, or touch a piece of cloth to determine its softness. There is a lot of interest in developing artificial systems that can recognize patterns, for example in automated speech recognition or visual object recognition. The following two sections describe related types of general pattern recognition systems: memory-based learning systems and Adaptive Resonance Theory neural networks.

2.2.1 Memory-based learning

The nearest neighbor (NN) algorithm (Cover and Hart, 1967; Dasarathy, 1991) is a well-known classification algorithm popular because it is very simple and it achieves high accuracy. In its basic version, the only learning step in the NN algorithm is to store all the training exemplars into its memory. After learning, when an unknown exemplar is presented, the NN algorithm assigns it to the class of its nearest neighbor among the stored exemplars. The simplest extension of the NN algorithm is the *k*-NN algorithm, in which, when an unknown exemplar is presented, multiple (*k*) nearest neighbors of the unknown exemplar are found and the class of the exemplar is determined by voting among the neighbors. The NN algorithm is the basic algorithm of a class of learning systems referred to as memory-based (Cybenko, Saarinen, Gray, Wu and Khrabrov, 1994), instance-based (Aha, Kibler and Albert, 1991), exemplar-based (Salzberg, 1990), case-based (Ram, 1993), experience-based (Sycara and Navinchandra, 1989), or lazy (Aha, 1997) learning algorithms. All these algorithms share the basic feature of the *k*-NN algorithm: the learning process consists of storing (a subset of) the training set into the memory; and classification of an unknown exemplar consists of finding the most similar stored exemplars and assigning the unknown exemplar to their class.

Well-known problems of the NN algorithm are its sensitivity to noise and large memory requirements. A majority of the memory-based learning systems are variations on the basic NN mechanism that try to eliminate these and other problems (see recent reviews by Dasarathy, Sánchez and Townsend, 2000; Wilson and Martinez, 2000; Lam, Keung and Liu, 2002). Several groups of memory-based learning algorithms can be defined depending on how the algorithms approach the following issues: direction of search (incremental or decremental), intuition about which exemplars to keep (center of clusters or border exemplars), distance function used, and feature weighting strategy.

The basic strategy to alleviate the large memory requirements of the memory-based learning systems is pruning (also called reduction or filtering) in which a system memorizes a subset of the original training set. The exemplars stored by a memory-based learning system are called nodes, and the set of stored nodes is called a coding set or code. Examples of pruning methods are the classical editing (Wilson, 1972) and condensing (Hart, 1968) rules for the NN algorithm. The pruning methods are usually simple and fast, however they might be unable to find the optimum solution if the best representatives are not included in the training set.

Direction of the search refers to the method the algorithm uses to generate the code. Incremental learning methods (e.g., the condensed nearest neighbor, CNN) (Hart,

1968) start with an empty coding set and sequentially add nodes into the code depending on whether they are considered to be useful. Decremental learning methods (e.g., the edited nearest neighbor, ENN) (Wilson, 1972) start by copying the whole training set into the code. Then, for each node in the code they evaluate whether it contributes to performance of the system, eliminating the nodes with the least contribution. There are several differences between the incremental and decremental pruning methods. Incremental methods are less computationally demanding and they require less memory. On the other hand the decremental methods are not sensitive to the training set ordering and the code they find is usually smaller and better than the code generated by the incremental systems.

Another factor that distinguishes memory-based learning systems is whether they prefer to retain the nodes that are near or far from the approximated decision boundary. The intuition behind retaining nodes near the boundary is that these nodes approximate the decision boundary while the "center" nodes influence the decision very rarely. On the other hand, retaining the center nodes can significantly reduce the code size and it can make the system resistant to noise while still correctly approximating the decision boundary. Some systems combine these strategies by looking for best center nodes and best border nodes separately, and then combining them (Lam et al., 2002).

Various distance metrics (also called similarity measures) can be used to determine the nearest neighbors of an exemplar. For continuous-valued features the most commonly used distance metrics are the Euclidian or the city-block metric or, more generally, $d(\vec{x},\vec{y}) = \left(\sum_{i=1}^{m} \left(\left|x_i - y_i\right|\right)^n\right)^{1/n}$, where $\vec{x}$ and $\vec{y}$ are vectors, m is the dimension of the input space, with n=1 for the city-block and n=2 for the Euclidian distance. Other continuous-value distance functions include the Minkowsky, Mahalanobis, Camberra, Chebychev, Quadratic, Correlation, or Chi-square distance metrics (Wilson and Martinez, 2000). Specific distance metrics were developed for features with nominal (discrete, unordered) values. The simplest one is the overlap metric, which is 0 if the two input vectors are equal and 1 otherwise. An alternative metric for nominal features is the Value Difference Metric (Stanfill and Waltz, 1986) which considers two nominal values to be closer if they have more similar classifications, regardless of their ordering. In addition, distance metrics have been developed that can handle both nominal and continuous features, as well as features with missing values (Wilson and Martinez, 1997).

If the input space is multidimensional, some features (dimensions) can be more informative than others and different dimensions can be contaminated by different amounts of noise. Therefore, it is important for memory-based learning systems to use a proper form of feature weighting that will stress the informative features and suppress the noisy ones. An extensive review of feature-weighting methods is available in Wettschereck, Aha and Mohiri (1997).

Memory-based learning systems are susceptible to various forms of the *curse of dimensionality*. For example, the search time needed to find the nearest neighbor grows exponentially with the number of input dimensions. Also, the number of training inputs needed to cover the input space with equal density grows exponentially with the input

space dimensionality. Several methods have been proposed to overcome this curse in these systems (Cybenko et al., 1994).

2.2.2 Adaptive Resonance Theory

Adaptive Resonance Theory (ART) was introduced by Grossberg (1976) as a theory of human cognitive information processing. Based on the theory, a series of real-time neural network architectures for unsupervised and supervised learning have been developed. These networks combine fast learning with stable category coding and are a suitable tool for many pattern recognition problems. The ART models for unsupervised learning include ART 1 (Carpenter and Grossberg, 1987a), ART 2 (Carpenter and Grossberg, 1987b), fuzzy ART (Carpenter, Grossberg and Rosen, 1991), and distributed ART (Carpenter, 1997). ARTMAP, a family of supervised ART architectures developed for classification problems, includes fuzzy ARTMAP (Carpenter et al., 1992), and distributed ARTMAP (Carpenter, Milenova and Noeske, 1998) neural networks. A collection of papers on ART models can be found in (Carpenter and Grossberg, 1991), more recent models are summarized in Carpenter et al. (1998).

ARTMAP neural networks share certain characteristics with the memory-based learning systems. Therefore, some characteristics of the memory-based learning systems described in the previous section apply also to the ARTMAP architectures.

Chapter 3 Spatial unmasking of nearby pure-tone sources

Abstract

Detection thresholds for 500-Hz and 1000-Hz pure-tone targets (T) were measured in the presence of a broadband masker (M) for different spatial configurations of T and M. Sources were simulated in anechoic space for source positions within reach of the listener, varying not only in azimuth (- 90° to +90° in 45° steps) but also distance (15 and 100 cm). For the spatial configurations tested, the T detection thresholds range over 50 dB (a much larger range than occurs when sources are more distant), primarily due to large energy effects. Inter-subject differences in the thresholds are large; however, the pattern of results is similar across subjects. For M at 0˚ or 45˚, the thresholds decrease with azimuthal separation of T and M and increase with T distance for both T frequencies. For M at 90˚, results are more complex. In some of these cases, azimuthal separation of T and M yields little change or even a small increase in the threshold and the pattern of results depends on T frequency. The amount of energy reaching the listener's ears from different T and M locations was calculated from the individually-measured head-related transfer functions (HRTFs) used in the simulations. The changes in the amount of energy due to changes in T and M location combined with predictions of binaural unmasking from a modified version of the Colburn (1977a) model capture general trends in the pattern of spatial unmasking. Individual differences in HRTFs (in both monaural and binaural acoustic cues) and in binaural sensitivity influence spatial unmasking, especially for sources within reach of the listener. However, even after accounting for inter-subject acoustic differences, small but consistent deviations between subject-specific model predictions and behavioral results remain. In addition, results suggest that individuals differ not only in their overall sensitivity to binaural cues (as assumed in the Colburn model), but also in how their binaural sensitivity varies with the spatial position of (and interaural differences in) M.

3.1 Introduction and background

When listening for a target sound (T) in the presence of a masking sound (M), a listener's ability to perceive T is influenced by the locations of T and M. In general, it is easier to detect or recognize T when it is spatially separated from M compared to when the T and M are at the same position. This "spatial unmasking" effect has been studied for many types of stimuli, including speech (e.g., see Freyman et al., 1999; Shinn-Cunningham, Schickler, Kopčo and Litovsky, 2001), click-trains (e.g., see Saberi et al., 1991; Good et al., 1997), and tone complexes (e.g., see Kidd et al., 1998).

For broadband noise maskers, spatial unmasking arises primarily from energetic and binaural effects. Energetic unmasking occurs because under many circumstances the target-to-masker ratio (TMR) increases at one ear when T and M are spatially separated compared to when they are at the same location. Binaural unmasking can occur when the interaural time and intensity differences caused by T and M differ.

There are many studies of how binaural differences influence tone detectability in noise (for a review of this classic literature, see Durlach and Colburn, 1978). However, most of these studies were performed under headphones using interaural differences that do not occur naturally. There are only a few studies that have measured how tone detection is affected by the spatial locations of T and M (examples include Ebata et al., 1968; Gatehouse, 1987; Santon, 1987; Doll and Hanna, 1995). Moreover, results of these studies are inconsistent, finding spatial unmasking ranging from as little as 7 dB (Santon, 1987) to as much as 24 dB (Gatehouse, 1987). These apparent discrepancies may be due to differences in the spatial configurations tested. However, none of these studies analyzed how T and M levels at the ears changed with spatial configuration and did not factor out how energetic and binaural factors may have contributed to the observed spatial unmasking.

Previous studies of spatial unmasking for pure tone targets considered sources relatively far from the listener and looked only at unmasking due to azimuthal separation, ignoring any effects of source distance. For sources more than about a meter from the listener, the only significant effect of changing source distance is a change in signal level that is equal at the two ears. However, changes in source distance for sources within reach of the listener produce changes in signal level that differ at the two ears, resulting in exceptionally large interaural level differences (ILDs; see Brungart and Rabinowitz, 1999; Shinn-Cunningham et al., 2000), even at low frequencies (for which ILDs are essentially nonexistent for relatively distant sources). In addition, for near sources, relatively small positional changes can lead to large changes in the energy of the T and M reaching the two ears.

A few previous studies suggest that binaural performance can be worse than monaural performance using the ear with the best TMR (the "better ear"), particularly when there are large ILDs in the stimuli (e.g., see Bronkhorst and Plomp, 1988; Shinn-Cunningham et al., 2001). Given that large ILDs can arise when sources are within reach of the listener, studies of binaural unmasking for nearby sound sources may shed light on these reports.

The current study examines spatial unmasking of pure tone sources within reach of a listener in a simulated anechoic environment. Individually-measured head-related

transfer functions (HRTFs) were used to simulate sources. This approach allowed realistic spatial acoustic cues to be presented to the subjects while still allowing detailed analyses of the stimuli reaching the subjects during the experiment. The main goals of the study are to: 1) measure how T threshold depends on T and M azimuth and distance for nearby sources, 2) characterize energetic effects by analyzing how the TMR varies with the spatial configurations tested, 3) evaluate the binaural contribution to spatial unmasking, particularly for spatial configurations in which large ILDs arise, and 4) investigate the degree to which results can be accounted for by a model of binaural interaction.

The current study begins by presentation of the results of a behavioral experiment that measured spatial unmasking of pure tone targets. The individualized HRTFs used to generate headphone stimuli in the experiment are then analyzed and compared to estimates from a spherical head model (Brungart and Rabinowitz, 1999; Shinn-Cunningham et al., 2000) and to HRTFs measured on a KEMAR acoustic manikin. The individualized HRTFs are then used to estimate the energetic and binaural contributions to spatial unmasking. Finally, binaural contributions to spatial unmasking, estimated by subtracting off energetic effects, are compared to predictions from the Colburn model of binaural processing (Colburn, 1977a; see also Stern and Shear, 1996) to evaluate whether the model can predict results for realistic sources near the listener.

3.2 Spatial unmasking of nearby pure tone targets

3.2.1 Methods

3.2.1.1 Subjects

Four graduate students with prior experience in psychoacoustic experiments (including author NK) participated in the study. One subject was female and three were male. Subject ages ranged from 25 - 28 years. All subjects had normal hearing as confirmed by an audiometric screening.

3.2.1.2 HRTF measurement

Individualized HRTF measurements were made with subjects seated in the center of a quiet classroom (rough dimensions of 5 x 9 x 3.5 meters; broadband T_{60} of approximately 700 ms). Subjects were seated with their heads in a headrest so that their ears were approximately 1.5 m above the floor. Measurements were taken for sources in the right front horizontal plane (at ear height) for all six combinations of azimuths (0º, 45º, 90º) and distances (0.15 m, 1 m), as shown in Figure 3-1.

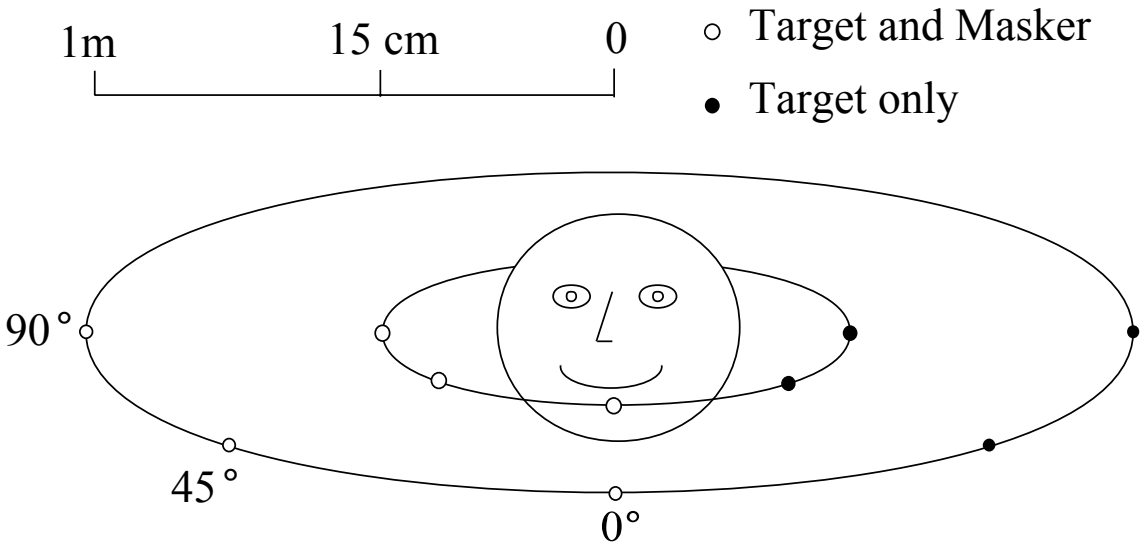


Figure 3-1 Spatial positions used in the study. HRTFs were measured at the positions denoted by open symbols. Target detection thresholds were measured for all spatial combination of six masker positions (open symbols) and ten target positions (filled and open symbols; targets simulated at the filled symbols used the corresponding HRTFs from the contralateral hemifield with left- and right-ear signals reversed).

The Maximum-Length-Sequence (MLS) technique (Vanderkooy, 1994) was used to measure HRTFs. Two identical 32,767-long maximum length sequences were concatenated and presented through a small loudspeaker using a 44.1 kHz sampling rate (details regarding the equipment are described below). The response to the second sequence was recorded. This measurement was repeated ten times in rapid succession and the raw measurements averaged in the time domain. This average response was cross-correlated with the original sequence to estimate a 743-ms-long impulse response. No correction for the measurement system transfer function was performed, but the amplitude spectrum of the transfer-function of this measurement system was examined and found to vary by less than 2 dB and to cause no significant interaural distortion for frequencies between 400 and 1500 Hz (the frequency region important for the current study). The useful dynamic range of the measurements (taking into account the ambient acoustic and electrical noise) was at least 50 dB for all frequencies greater than 300 Hz.

HRTFs were measured using a Tucker-Davis Technologies (TDT) signal processing system under computer control. For each measurement, the concatenated MLS sequence was read from the PC hard-drive and sent to a TDT D/A converter (TDT PD1), which drove an amplifier (TDT HB6) connected to a BOSE mini-cube loudspeaker. A Polhemus FastTrak electromagnetic tracker was used to ensure that the subject's head was positioned correctly relative to the loudspeaker (which was hand-positioned by the experimenter prior to each measurement). Miniature microphones (Knowles FG-3329c) mounted in earplugs and inserted into the entrance of the subjects' ear canals (to produce blocked-meatus HRTF recordings) measured the raw acoustic responses to the MLS

sequence. Microphone outputs drove a custom-built microphone amplifier that was connected to a TDT A/D converter (TDT PD1). These raw results were stored in digital form on the computer hard-drive for off-line processing to produce the estimated head-related impulse responses.

HRTFs measured as described above include room echoes and reverberation. To eliminate room effects, time-domain impulse responses were multiplied by a 6-ms-long $\cos^2$ time window (rise/fall time of 1 ms) to exclude all of the reverberant energy while retaining all of the direct-sound energy. The resulting "pseudo-anechoic" HRTFs were used to simulate sources (and in all subsequent analysis).

HRTFs were measured only for sources in the right hemifield. To simulate sources in the left hemifield, HRTFs from the corresponding right-hemifield position were used, exchanging the left and right channels (i.e., left/right symmetry was assumed; given that only pure tone targets were simulated in the left hemifield, this approximation should introduce no significant perceptual artifacts in the simulated stimuli).

3.2.1.3 Stimulus generation

Target stimuli consisted of 165-ms-long pure tones (500 or 1000 Hz) gated on and off by 30-ms $\cos^2$ ramps. The T was temporally centered within a broadband, 250-ms-long masker. On each trial, the M token was randomly chosen from a set of 100 pre-generated samples of broadband noise that were digitally low-pass filtered with a 5000 Hz cutoff frequency (9^{th} order Butterworth filter, as implemented in the signal-processing toolbox in Matlab, produced by the Mathworks, Natick, MA).

T and M were simulated as arising from different locations in anechoic space by convolving the stimuli with appropriate individualized HRTFs. The simulated spatial configurations included all combinations of T at azimuths (-90°, -45°, 0°, 45°, 90º) and distances (0.15 m, 1 m) and M at azimuths (0°, 45°, 90˚) and distances (0.15 m, 1 m). A total of 60 spatial configurations were tested (10 T locations x 6 M locations; see Figure 3-1).

For nearby sources, keeping the M presentation level constant would result in the received level (at the subject's ears) varying widely with M position. In order to keep the received level of M relatively constant, the levels of the HRTF-processed M stimuli were normalized to fix the RMS energy falling within the equivalent rectangular band (ERB; Moore, 1997) centered on the T frequency at the ear receiving the more intense M signal (the right ear for all of the tested configurations). For the 500-Hz center frequency, a 100 Hz ERB was used. For the 1000-Hz target, the ERB width was set to 136 Hz. The right- (louder-) ear RMS masker level in the ERB was fixed at 64 dB SPL (the values used to equalize the masker for each M location are shown in Figure 3-4). The amounts by which M was normalized were added back to the raw measured thresholds to predict the amount of spatial unmasking that would occur if the signal level emitted by a free-field M were held constant (note that this analysis assumes that detection performance depends only on the target to masker ratio or TMR and is independent of the overall M level).

Stimulus files, generated at sampling rate of 44.1 kHz, were stored on the hard disk of the control computer (IBM PC compatible). In each trial, appropriate T and M samples were presented through Tucker-Davis Technologies (TDT) hardware. Left- and

right-ear T and M signals were processed through four separate D/A converters (TDT PD1). Target signals were scaled to the appropriate presentation level by a programmable attenuator (TDT PA4), summed with the masker signals (TDT SM3), and amplified through a headphone buffer (TDT HB6). The resulting binaural stimuli were presented via Etymotic Research ER-1 insert earphones. No filtering was done to compensate for the transfer characteristics of the playback system. A handheld RS 232 terminal (QTERM) was used to gather subject responses and provide feedback.

3.2.1.4 Experimental procedure

Behavioral experiments were performed in a single-walled sound-treated booth.

A 3-down-1-up, three-interval, two-alternative, forced-choice procedure was used to estimate detection thresholds (Levitt, 1971), defined as the 79.4% correct point on the psychometric function. Each run started with the T at a clearly detectable level and continued until 11 "reversals" occurred. The T level was changed by 4 dB on the first reversal, 2 dB on the second reversal, and 1 dB on all subsequent reversals. For each adaptive run, detection threshold was estimated by taking the average T presentation level over the last six reversals. At least three separate runs were performed for each subject in each condition. Final threshold estimates were computed by taking the average threshold across the repeated adaptive threshold estimates. Additional adaptive runs were performed as needed for every subject and condition to ensure that the standard error in this final threshold estimate was less than or equal to 1 dB.

The study was divided into two parts, one measuring thresholds for the 500-Hz T and one for the 1000-Hz T. Three subjects performed each part (two of the four subjects performed both). For each target, subjects performed multiple sessions consisting of ten runs. Subjects were allowed to take short breaks between runs within one session, with a minimum 4-hour break required between sessions. Each subject performed one initial practice session consisting of four practice runs and six runs measuring detection thresholds for NoSo and NoSπ conditions (where NoSo represents a sinusoidal diotic signal, i.e., with zero interaural phase, in the presence of a diotic noise; NoSπ represents a sinusoidal signal with interaural phase difference equal to π in the presence of a diotic noise). Subjects then performed eighteen additional sessions (180 runs; 3 runs each of every combination for 6 T positions and 10 M positions). In each of these sessions, a full set of thresholds was determined for one M position (the order of the 10 T positions was randomized within each session). These sessions were grouped into three blocks of six, each block containing a full set of thresholds. The order of M positions was separately randomized for each block and subject. Each subject performed approximately 20 hours of testing per T frequency.

3.2.2 Results

3.2.2.1 Binaural masking level difference

Table 3-1 shows the binaural masking level difference (BMLD; see Durlach and Colburn, 1978), defined as the difference in T detection threshold in the NoSo and NoS□

conditions. As expected from results of previous studies, BMLDs are larger for the 500-Hz T (where BMLDs ranged from 11 – 16 dB) than the 1000-Hz T (where BMLDs ranged from 7 – 14 dB). Also consistent with previous reports, inter-subject differences in the BMLD are large and fairly consistent across T frequency. For instance, Subject S1 has the largest BMLDs at both T frequencies whereas Subject S2 has the smallest BMLDs at both frequencies.

3.2.2.2 Spatial unmasking

The amount of "spatial unmasking" is defined as the change in the energy that T would have to emit to be just detectable when at the simulated location compared to the energy emitted by a just-detectable T located at the same position as M. In order to estimate the T detection threshold when the emitted level of M is held constant, the amount by which the M was normalized was first added back to the raw T detection thresholds. Then, an average of the thresholds when T and M are at the same location was computed, and this value was subtracted from all the thresholds to obtain an estimate of spatial unmasking (i.e., the amount by which detection thresholds improve with spatial separation of T and M).

	Measured BMLD (dB)				
	Individual Subjects				Subject
	S1 (●)	S2 (■)	S3 (▼)	S4 (▲)	Average
500 Hz	15.6	11.0	14.5		13.7
1000 Hz	13.1	7.5		8.7	9.8

Table 3-1 Binaural masking level differences

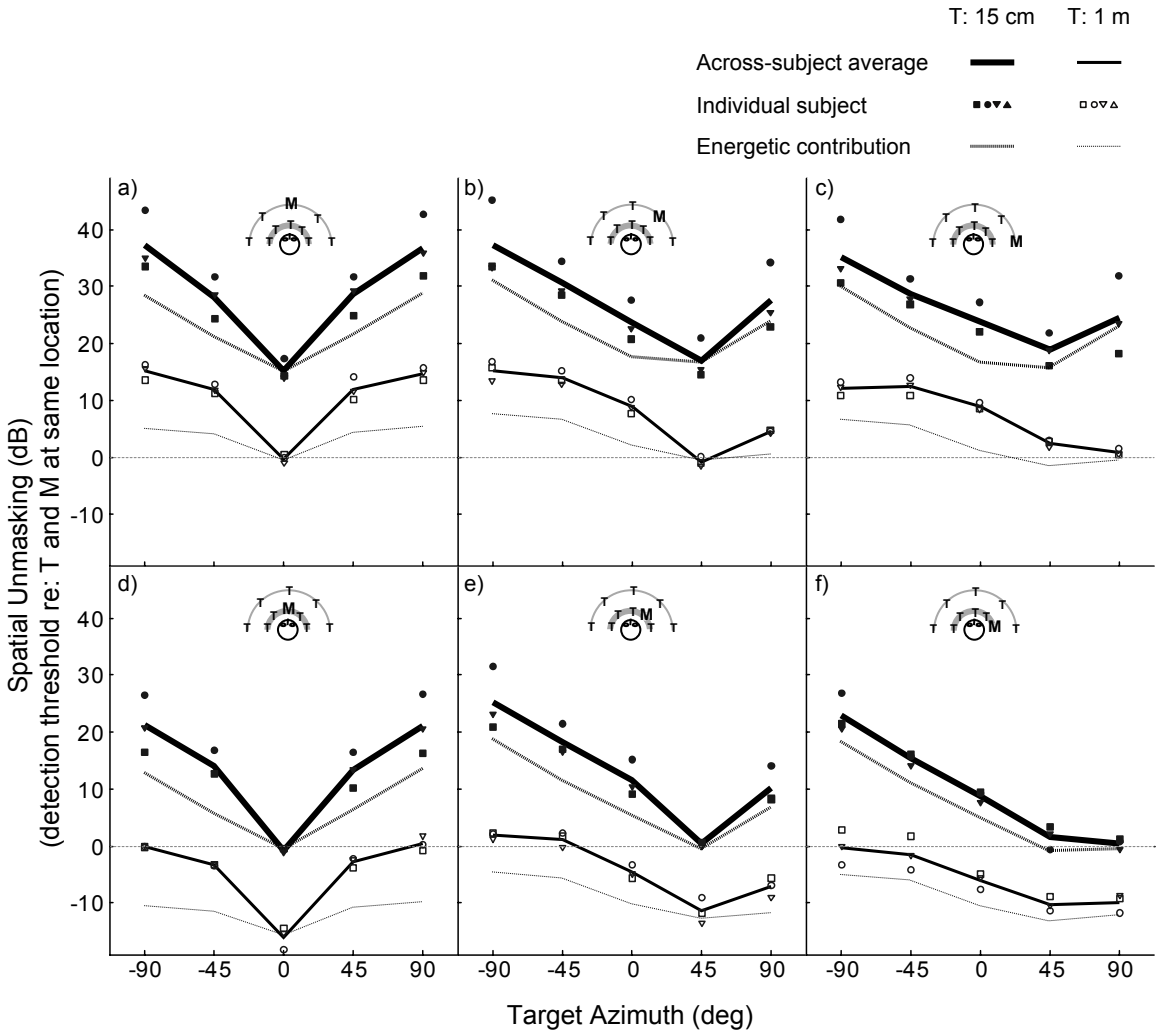


Figure 3-2 Spatial unmasking for the 500-Hz T. Each panel plots spatial unmasking (the difference between target detection threshold when T and M are at the same spatial location and when T and M are in the spatial configuration denoted in the plot) as a function of T azimuth for a fixed M location. Across subject averages are plotted for T distances of 15 cm (thick solid lines) and 1 m (thin solid lines). Individual subject results are plotted as symbols. Dashed lines show the estimated energetic contribution to spatial unmasking. The spatial configurations of T and M represented in each panel are denoted in the panel legend.

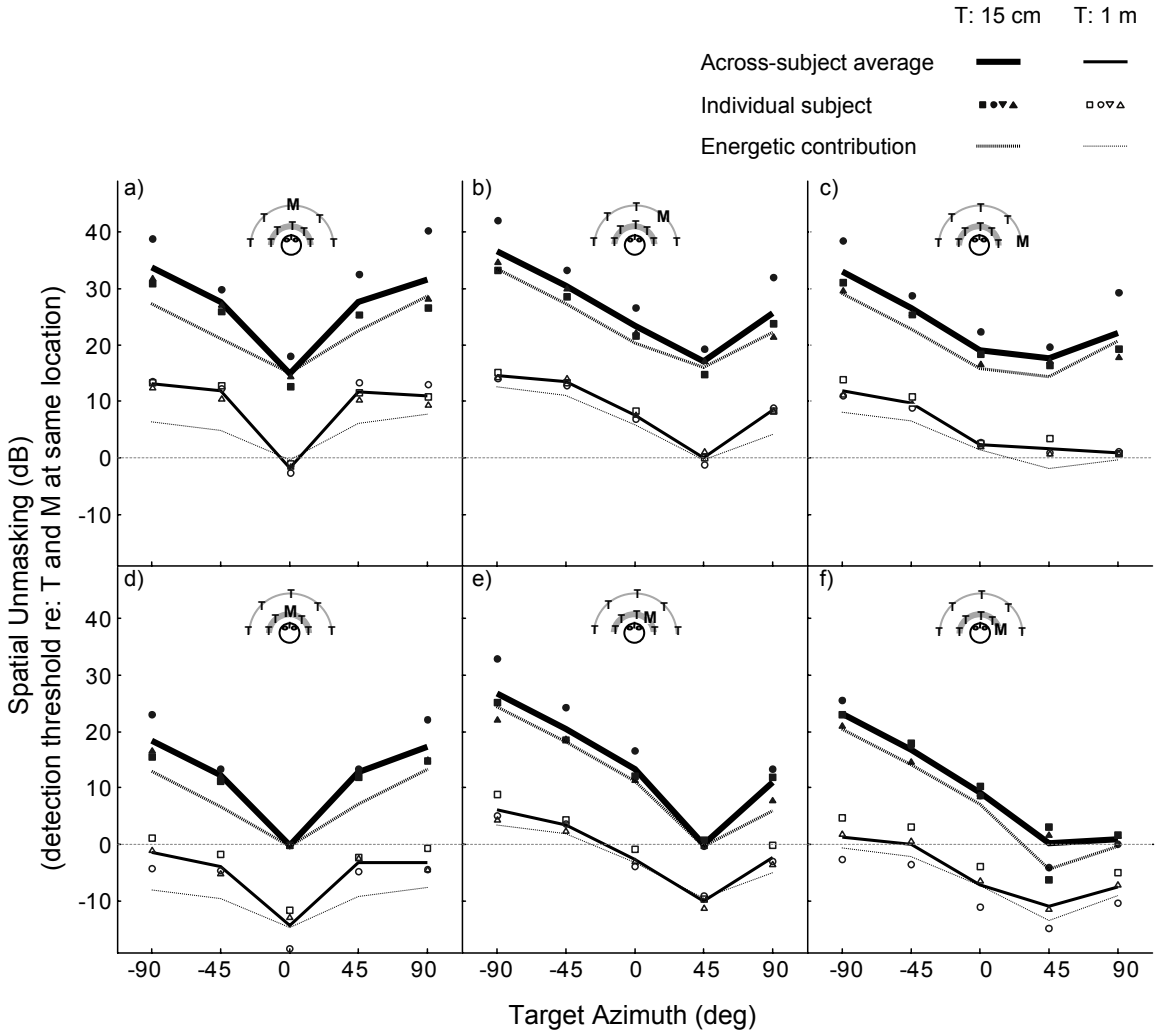


Figure 3-3 Spatial unmasking for the1000-Hz T. See caption for Figure 3-2.

Figure 3-2 and Figure 3-3 plot the amount of spatial unmasking of 500- and 1000-Hz targets, respectively, for the across-subject average (solid lines) and the individual subjects (symbols. Dashed lines show predicted energetic effects at the better ear discussed in Section 3.4). Individual subject results were calculated by averaging the adaptive-run threshold estimates over all repetitions for each spatial configuration. Across-subject averages were calculated by averaging these individual-subject averages.

For the spatial configurations tested, the amount of spatial unmasking spans a range of over 50 dB, with larger unmasking observed at 500 Hz than at 1 kHz. While subjects generally show similar patterns of results, inter-subject differences are large. For instance, Subject S1 shows as much as 10 dB more unmasking than the other subjects [e.g., when T is at (90˚, 15 cm) and M is at (0˚, 1 m), Figure 3-2 and Figure 3-3, thick lines in panel a]. However, this same subject consistently shows the least unmasking when M is at (90˚, 15 cm) and T is at 1 m (thin lines in panel f).

Despite the large inter-subject differences, overall trends are similar across subjects and for both 500- and 1000-Hz targets. Unsurprisingly, for both target frequencies, positioning T near to the subject (thick solid line in each panel of Figure 3-2 and Figure 3-3) improves T detectability compared to when T is far from the subject (thin solid line in each panel). Similarly, positioning M near the subject (lower panels) degrades T detectability compared to when M is farther from the subject (upper panels). For sources within reach, examined in this experiment, the source distance matters even when both T and M are at the same distance. Specifically, the amount of unmasking due to angular separation of T and M increases with decreasing distance (thin lines in panels a,b,c show less unmasking than thick lines in panels d,e,f).

To a first order approximation, changes in M distance cause a simple shift in the amount of spatial unmasking (i.e., results in the upper panels are essentially the same as results in the lower panels shifted upward by 10-15 dB). However, closer inspection shows that the interaction of azimuth and distance is more complex. For example, when M is at 45˚ or 90˚ and T is at 15 cm, the change in spatial unmasking with T angle is greater when M is at 15 cm than when M is at 1 m (thick lines in panels b,c vs. e,f). Changing T distance also causes complex effects; the amount of spatial unmasking varies more with T angle when T is close to the listener (thick lines) compared to when T is at the farther distance (thin lines).

In general, separating T and M in azimuth improves T detectability compared to when T and M are in the same direction, independent of whether T and M distances are the same or different. For instance, when M is at 0˚ the lowest points in each plot arise when T is at 0˚ (leftmost panels in Figure 3-2 and Figure 3-3); when M is at 45˚ the lowest points arise when T is at 45˚ (middle panels). However, when M is at 90˚, angular separation of T and M does not always increase the amount of unmasking. For instance, for both the 500-Hz and 1000-Hz T with M at (90˚, 1 m), there is less spatial unmasking when the 15-cm T is at 45˚ than when it is at 90˚ (Figure 3-2 and Figure 3-3, thick line in panel c). Similarly, when M is at (90˚, 15 cm) and T is at 1 m, the amount of unmasking is either equal for T at 45˚ and 90˚ (500-Hz T: Figure 3-2, thin line in panel f) or greater when T is at 90˚ compared to when T is at 45˚ (1000-Hz T: Figure 3-3, thin line in panel f).

3.2.3 Discussion

Both the size of the BMLD and the amount of spatial unmasking varies from subject to subject. Differences in spatial unmasking may be partially explained by the inter-subject differences in the size of the BMLD. For instance, Subject S1 has the largest BMLDs and exhibits the most spatial unmasking. However, differences in spatial unmasking could also be due to differences in the acoustic parameters in the individually-measured HRTFs. Acoustic differences in the measurements and the binaural contribution to spatial unmasking are considered further in Section 3.4. Taken together, these analyses suggest that inter-subject differences in spatial unmasking are affected both by differences in acoustic cues and in different sensitivities to binaural cues.

Many of the current results follow easily predicted patterns. Moving T closer to the subject improves detection performance (as expected on the basis of an increase in the

level of T reaching the listener); conversely, moving M closer degrades detection performance (as expected when the level of M at the ears increases). Separating T and M in angle improves detection performance for most spatial configurations, and the amount of this improvement increases as the sources are moved closer o the head. However, there are other effects that are less intuitive. Unmasking varies less with T azimuth for a 15-cm M than for a 1-m M. For the same angular separation of T and M, unmasking decreases with the laterality of M. Finally, when T and M are at different distances and M is at 90˚, the amount of unmasking can actually decrease when T is at 45˚.

Apparent discrepancies in the amount of spatial unmasking due to angular separation of T and M observed in previous studies are actually consistent with the current results. For example, the current study found more spatial unmasking for 1 m sources when M is at 0˚ (Figure 3-2 and Figure 3-3, panels a and d) than when M is at 90˚ (panels c and f). Thus, the relatively large amount of spatial unmasking observed by Gatehouse (1987) compared to that found by Santon (1987) may be due to the fact that Gatehouse fixed M in front of the listener and varied T azimuth, whereas Santon fixed T in front of the listener and varied M azimuth. In other words, all of these results are consistent with the idea that the amount of spatial unmasking is larger when M is in front of the listener and T is displaced laterally than when T is in front of the listener and M is angularly displaced.

3.3 HRTF measurements

The acoustic factors that influence spatial unmasking can be characterized by analysis of the HRTFs used in the simulations. Three acoustic characteristics of the HRTFs influence the performance in a spatial unmasking task: the magnitude spectra of, the interaural level differences (ILDs) in, and the interaural time differences (ITDs) in the signals reaching the two ears. The magnitude spectra of the HRTFs determine the intensity of the sound at the ears and thus the amount of spatial unmasking due to energy effects. ITDs and ILDs determine the amount of binaural unmasking. In this section, these parameters are analyzed for the individualized HRTFs used in the previous experiment.

HRTFs from the individualized HRTFs are also compared to values measured for a KEMAR acoustic manikin and predicted by a spherical model of the head. While the literature contains descriptions of both KEMAR (Brungart and Rabinowitz, 1999) and spherical-head model (Duda and Martens, 1998; Shinn-Cunningham et al., 2000) HRTFs for sources near the listener, the current analysis compares these “generic” models to HRTFs from human subjects to determine whether the models capture the acoustic effects that are important for predicting the amount of spatial unmasking as a function of nearby T and M locations.

3.3.1 Methods

KEMAR HRTFs were measured using a procedure identical to that used for the human listeners (see description in Section 3.2). HRTF predictions for a spherical head model (Duda and Martens, 1998; Brungart and Rabinowitz, 1999; Shinn-Cunningham et al., 2000) were computed using a head with radius of 9 cm and diametrically-opposed

ears. These results are compared to the HRTFs measured for the four subjects who participated in the spatial unmasking experiment.

For all of the HRTFs, the magnitude spectra, ILD, and ITD were determined for the equivalent rectangular band (ERB) centered at a given frequency. Magnitude spectra were calculated as the RMS energy in the HRTF falling within each ERB filter (100-Hz width centered at 500 Hz and 136-Hz width centered at 1000 Hz). ILDs were computed as the difference in the magnitude spectra for the left and right ears. ITD was first estimated as a function of frequency by taking the difference between the right- and left-ear HRTF phase angles at each frequency f and dividing by $2\pi f$. The ITD in each ERB filter was then estimated as the average of the ITD values for the frequencies falling within each ERB filter.

3.3.2 Results

3.3.2.1 Energetic effects

Figure 3-4 shows the magnitude of the ERB-filtered HRTFs at 500 and 1000 Hz for the left ear relative to a source at (0˚, 1 m). (Recall that HRTFs were measured only for sources to the right of the listener and that this analysis assumes left-right symmetry.) Results are shown for individual human subjects (symbols), the across-human-subject average (solid line), KEMAR (dotted line), and a spherical head model (dashed line).

Unsurprisingly, for both frequencies the spectral gain increases with decreasing distance. However, in addition to an overall shift in level, the dependence of the HTRF level on source azimuth differs for the two distances. Specifically, for the 15 cm distance, the gain to the ipsilateral ear grows rapidly with source eccentricity compared to the 1 m distance (the gain to the contralateral ear changes similarly with source angle for both distances).

Overall, inter-subject differences are modest for the more distant source (lower plots in each panel). However, there are large inter-subject differences for the 15 cm source positions. For instance, for both of the analyzed frequencies the 15-cm HRTF gain for Subject S1 is 5-10 dB larger than for the other subjects, except at 45˚, where it is comparable to that of other subjects.

KEMAR measurements and spherical-head predictions are similar to the measurements taken on the human subjects: both KEMAR and spherical-head results generally fall within the range of values observed for the four human subjects. However, there are a few source positions for which the KEMAR and spherical-head results deviate from the human measurements. KEMAR and spherical head results generally overestimate the HRTF gain for a source at 45 and underestimate the gain at 1000-Hz when the sound source is ipsilateral to the ear being analyzed.

While intuitively we expect the level of the signal reaching the ears to vary monotonically with lateral angle of the source, human HRTF measurements show that this is not strictly true. In particular, the 1000-Hz human measurements show that less energy reaches the contralateral ear when a source is at 45˚ than when it is at 90˚. Similarly, at 500 Hz the gain to the contralateral ear is comparable for 45˚ and 90˚ sources rather than decreasing for the 90˚ source. This nonmonotonicity, which is

likely due in part to the acoustic "bright spot" (e.g., see Brungart and Rabinowitz, 1999), is underestimated in both the spherical-head model and KEMAR HRTFs, especially at 1000 Hz.

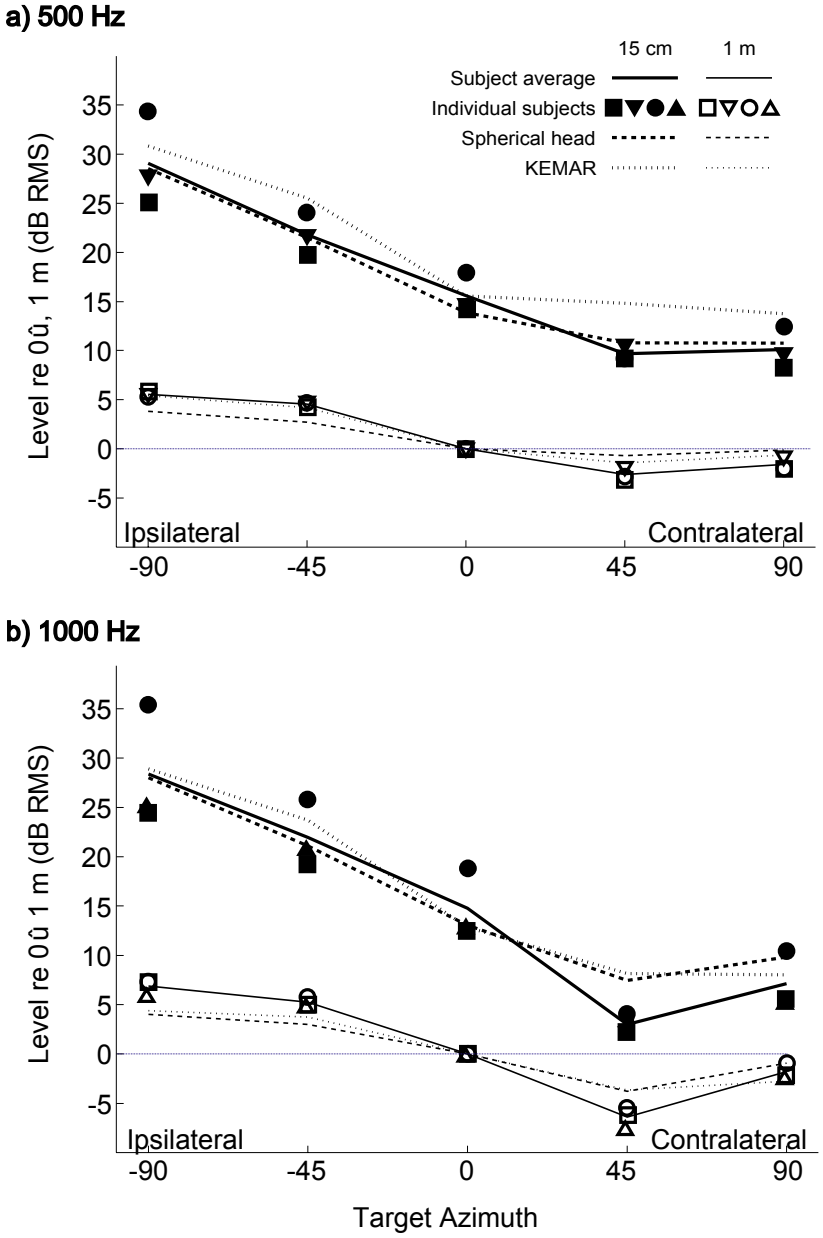


Figure 3-4 Left-ear HRTF spectrum levels in ERB filters, relative to the left-ear HRTF for a source at (0˚, 1 m). Results are shown for individual listeners, KEMAR, and the spherical head model as a function of source position. a) 500 Hz. b) 1000 Hz.

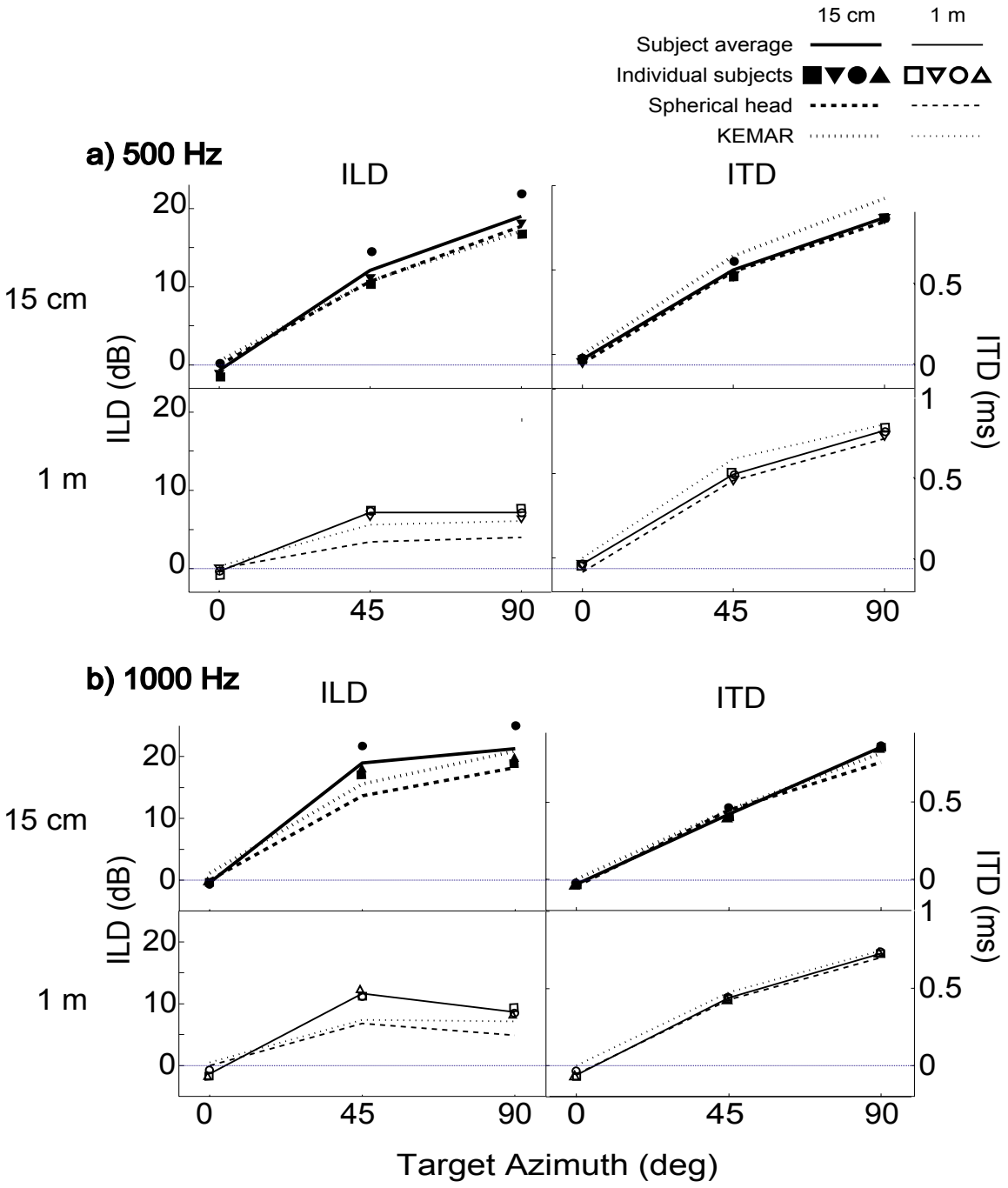


Figure 3-5 ILDs and ITDs in HRTFs for individual subjects, KEMAR manikin, and the spherical head model. a) 500 Hz. b) 1000 Hz.

3.3.2.2 Interaural differences

Figure 3-5 shows the ILDs and ITDs in the measured HRTFs at 500 and 1000 Hz (panels a and b, respectively) for the spatial positions used in the study. As in Figure 3-4, results for individual subjects (symbols), the across-human-subject average (solid line), KEMAR (dotted line), and a spherical head model (dashed line) are shown.

ILDs (left side of figure) were calculated directly from the measurements plotted in Figure 3-4. As a result, there are large inter-subject differences in the ILDs that are directly related to the inter-subject differences in the monaural spectral gains; for instance, Subject S1 has much larger ILDs for the 15 cm source than any of the other subjects.

As expected, ILDs are much larger for sources at 15 cm compared to 1 m, with ILDs at 500 and 1000 Hz approaching 20 dB for the nearby sources at 90˚. The spherical-head and KEMAR results tend to underestimate ILDs, although for the 500-Hz, 15-cm sources, both spherical-head and KEMAR results are within the range of human-subject observations. Discrepancies are most pronounced for a 1000-Hz source at a distance of 1 m and are greater for the spherical-head predictions than KEMAR measurements.

ITDs (right side of figure) vary primarily with source angle and change only slightly with distance and frequency. For most of the measured locations, both spherical-head and KEMAR results are in close agreement with human measurements. The only discrepancy between human results and model results is observed in the 500-Hz data, where the KEMAR measurements tend to have larger ITDs than the values measured in the human HRTFs.

3.3.3 Discussion

Overall, both spherical-head and KEMAR HRTFs provide reasonable approximations for how acoustic parameters in human HRTFs vary with source location. However, these models of human HRTFs produce small but consistent prediction errors (e.g., overestimating the gain at the contralateral ear when a source is at 45˚; underestimating the ILD for sources off midline, particularly at the 1 m distance).

Inter-subject differences in the HRTFs are large, especially for nearby sources. Of the four subjects, one subject showed consistently larger spectral gains and consistently larger ILDs than the other subjects when the source was at 15 cm. While it is possible that some of the inter-subject differences arise due to inaccuracies in HRTF measurement (e.g., due to hand-positioning the loudspeaker), the fact that one subject has consistently larger gains and ILDs for all nearby source locations suggests that real anatomical differences rather than measurement errors are responsible for the observed effects. It is also interesting to note that the observed inter-subject differences are much smaller for the 1 m source, suggesting that inter-subject differences in HRTFs are especially important when considering sources very close to the listener.

3.4 Energetic and binaural contributions to spatial unmasking

3.4.1 Analysis

For each subject, estimates of the energetic and binaural contributions to spatial unmasking were derived from the acoustic parameters of the HRTFs and the behavioral thresholds.

The energetic contribution to spatial unmasking was estimated by calculating the TMR at the better ear for each spatial configuration when T and M emit the same level (and thus would cause a TMR of zero when at the same location). The resulting TMR predicts the amount by which T thresholds decrease due to energetic effects at the better ear (i.e., the TMR is the same magnitude as but opposite in sign to the expected contribution of energetic effects to spatial unmasking). Across-subject average predictions were computed by averaging these estimates across subject. The binaural contribution to spatial unmasking was estimated for each subject by subtracting the estimated energetic contribution to spatial unmasking (derived from individually-measured HRTFs) from the individual behavioral estimates of spatial unmasking.

3.4.2 Results

3.4.2.1 Energetic contributions to spatial unmasking

While inter-subject differences in the energetic contribution to spatial unmasking are large, the trends in the across-subject average data capture the important features of the individual data. For brevity, only the across-subject averages are presented in Figure 3-2 and Figure 3-3 (for the 500- and 1000-Hz T, respectively, shown by dashed lines). For all spatial configurations tested, the behaviorally-observed amount of spatial unmasking either equals or is larger than the predicted spatial unmasking due to energy effects. Thus, even when there are large ILDs in the signals reaching the listener, binaural performance is always better than or equal to predicted performance when listening monaurally with the acoustically-better ear.

Better-ear energetic effects account for a large portion of the observed spatial unmasking when T and M are in the same direction and for the large influence of T and/or M distance on spatial unmasking at all T/M configurations. Generally, angular separation of T and M increases the energetic contribution to unmasking. However, when M is at 90˚, energetic effects either decrease or are roughly the same when T is at 45˚ compared to 90˚. Energetic contributions to unmasking change more with target azimuth when T is at 15 cm than at 1 m, primarily due to the fact that for nearby sources, small positional changes cause large changes in the relative distance from source to the ear.

Finally, energetic contributions to unmasking are relatively more important (i.e., account for a greater percentage of the observed amount of spatial unmasking) for the 1000-Hz T than the 500-Hz T. This is true both because the energetic effects are somewhat larger and because the additional spatial unmasking for which energy effects cannot account is smaller at 1000 Hz compared to 500 Hz.

Figure 3-6 Estimated binaural contribution to spatial unmasking for the 500-Hz T. Each panel plots the amount of binaural unmasking for one M position for both the 15-cm and 1-m T. Symbols show estimates for individual subjects with error bars showing the range of results across multiple adaptive runs. Lines trace a 2-dB range around the predicted amount of binaural unmasking from the Colburn (1977a) model for the 15-cm (dashed black lines) and 1-m (solid gray lines) T. The layout of the spatial configurations of T and M represented in each panel are shown in the Figure legend. a) Subject S1 b) Subject S2. c) Subject S3.

3.4.2.2 Binaural contributions to spatial unmasking

Figure 3-6 and Figure 3-7 show the estimated binaural contribution to spatial unmasking for the 500-Hz and 1000-Hz T. Plots show individual-subject estimates derived from behavioral data and HRTF analyses as symbols (the lines plot model predictions, derived and discussed in Section 3.5).

Even though inter-subject differences are large, there are a number of trends that are consistent across subjects. Unsurprisingly, there is no binaural unmasking when T and M are at the same spatial location. In fact, only Subject S1 shows any binaural unmasking (and only for the 500 Hz T) when T and M are at the same off-median-plane direction but at different distances (Figure 3-6 panel a, M at 45°, 15 cm). Overall, T distance has relatively little impact on results (compare circles and squares within each panel). However, M distance does influence results (compare upper panels and lower panels): binaural unmasking decreases when M is at 15 cm compared to 1 m, particularly for the 500-Hz results when M is located at 90˚.

The binaural contribution to spatial unmasking is larger for the 500-Hz T than the 1000-Hz T. For both T frequencies, the amount of binaural unmasking tends to be largest when M is at 0˚ (left panels) and decrease as M is displaced laterally (middle and right panels). Consistent with this observation, the change in binaural unmasking with T angle is smaller when M is laterally displaced than when M is at 0˚, particularly for the 1000-Hz T. For instance, for Subject S1 binaural contributions to spatial unmasking for the 1000-Hz T range from 0-8 dB when M is at 0˚ (depending on T azimuth); however, when M is at 90˚, binaural unmasking is roughly constant, independent of T angle (roughly 0-2 dB for the 15-cm M; roughly 2-4 for the1-m M).

The angular separation of T and M that leads to the greatest amount of binaural unmasking depends on T frequency. For the 500-Hz T, binaural unmasking tends to be greatest when T and M angles differ by about 90˚; however, for the 1000-Hz T, binaural unmasking tends to be greatest when T and M angles differ by roughly 45˚.

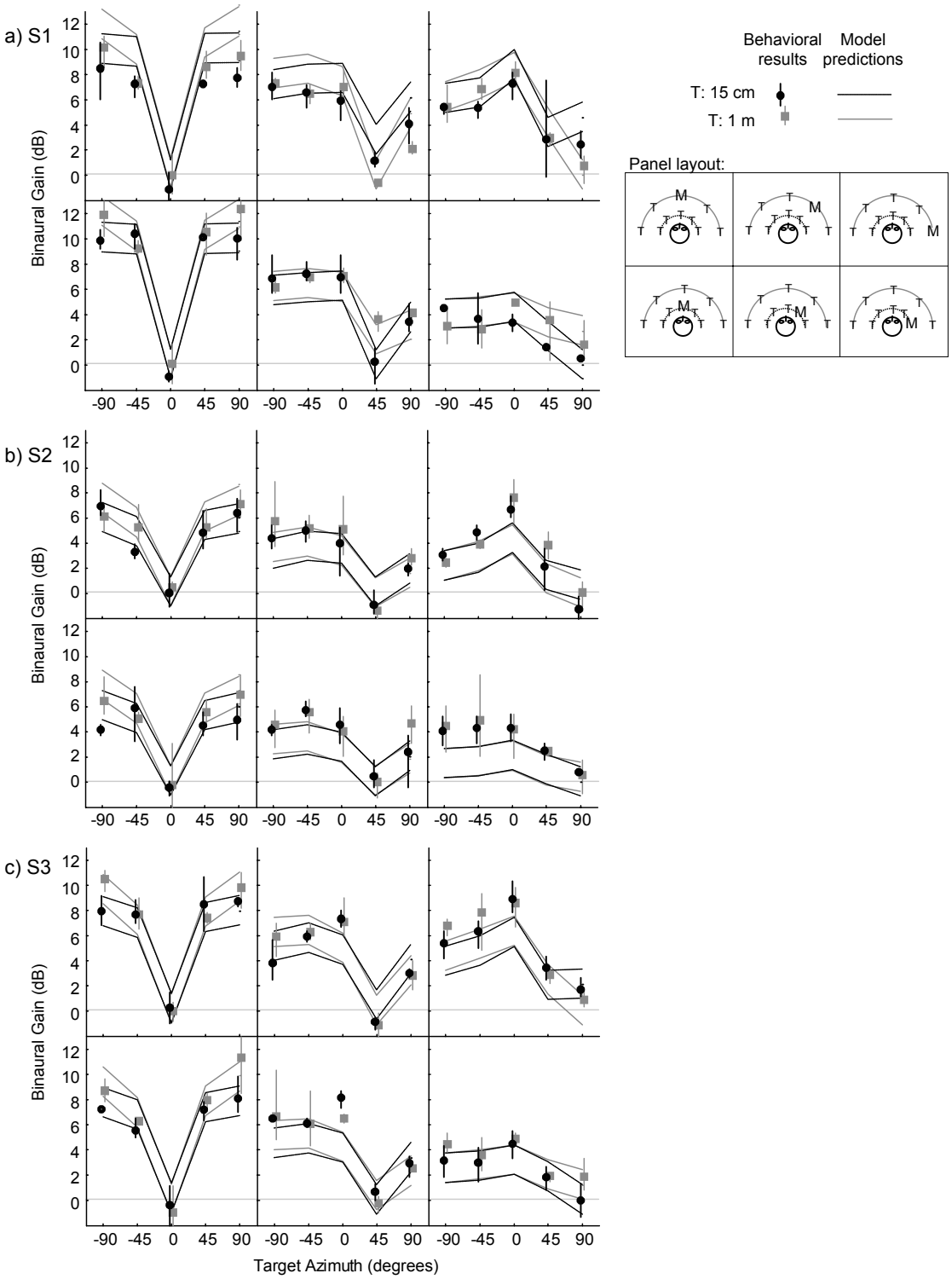


Figure 3-6 (caption on previous page)

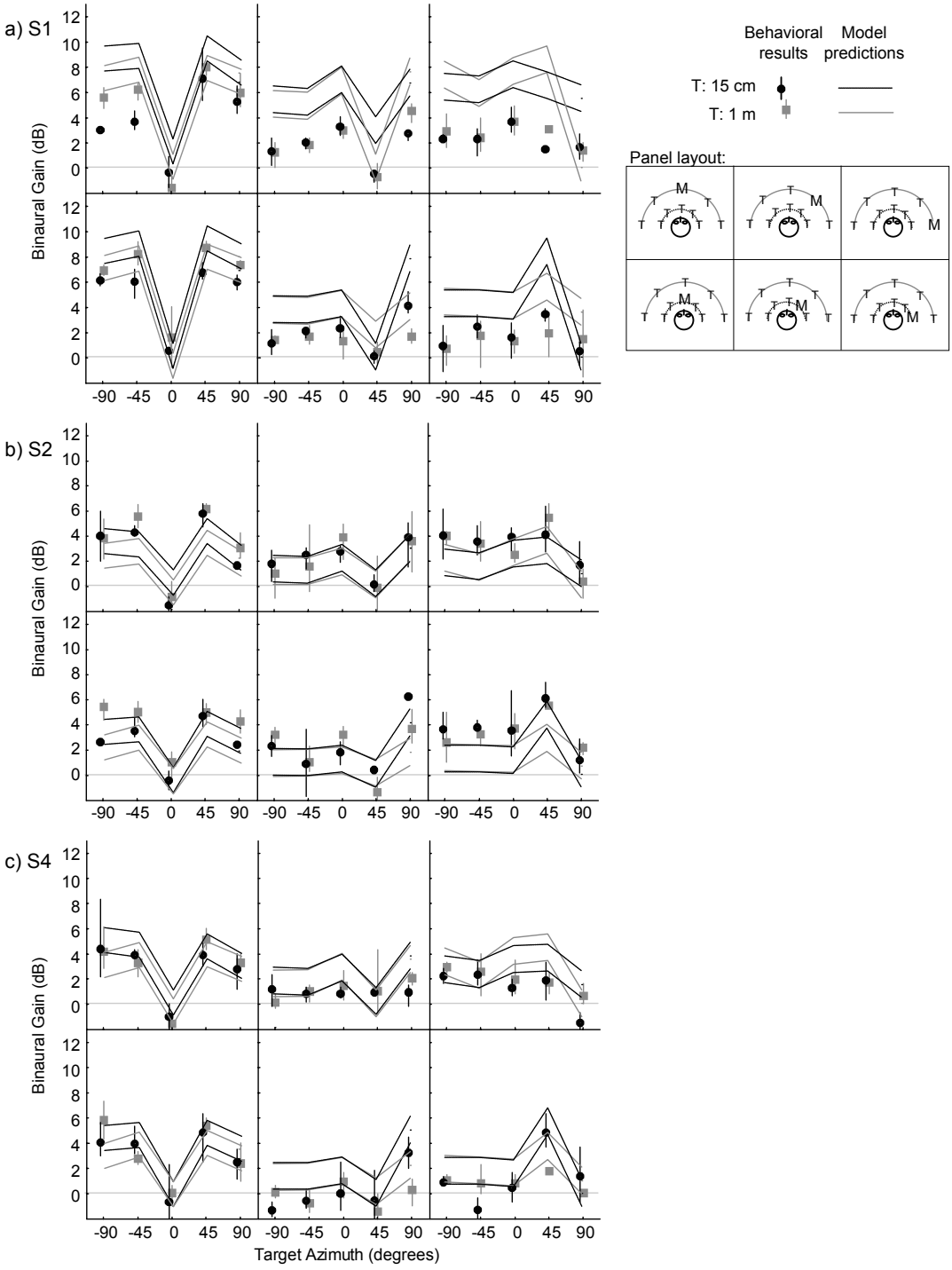


Figure 3-7 Estimated binaural contribution to spatial unmasking for the 1000-Hz T. See caption for Figure 3-6. a) Subject S1 b) Subject S2. c) Subject S4.

3.4.3 Discussion

Energetic factors contribute significantly to spatial unmasking for all of the spatial configurations tested. Energetic effects are larger at 1000 Hz than 500 Hz and are larger when T is at 15 cm compared to when T is at 1 m. The energetic contribution to spatial unmasking does not always increase monotonically with angular separation of T and M; in particular, when M is at 90˚, displacing T toward the median plane can lead to decreases in the TMR at the better ear, especially if T and M are at different distances. This result that helps explain why angular separation of T and M does not always improve detection performance.

Subjects show large differences in their ability to use binaural cues in detection tasks. For Subject S1, binaural differences can decrease detection thresholds by as much as 12 dB at 500 Hz, while for Subject S2 they provide at most 7 dB of unmasking. These differences in spatial unmasking roughly correlate with differences in BMLDs (Table 3-1); however, inter-subject differences in binaural sensitivity for one masker location do not predict results in other spatial configurations. For example, Subjects S1 and S3 show much more unmasking than Subject S2 in the 500-Hz conditions when M is at 0˚. However, when M is at 90˚, all three subjects exhibit essentially the same amount of binaural unmasking. This result suggests that inter-subject differences in binaural sensitivity cannot be fully captured with a single "binaural sensitivity" parameter at each frequency (the degree to which inter-subject differences can be predicted by the Colburn, 1977a model is considered further in Section 3.5). Instead, it seems that the distribution of the binaurally sensitive units as a function of the ITD and/or ILD might differ from subject to subject.

The magnitude of interaural level differences in M appears to have a large effect on the amount of binaural masking. Binaural unmasking is greatest when M is at 0˚ (and ITDs and ILDs in M are near zero); when M is at 45˚ and 90˚, the amount of binaural unmasking decreases. When M is off to the side, the binaural contribution to spatial unmasking is also smaller when M is at 15 cm compared to 1 m. These effects are consistent with past reports showing that the BMLD decreases with masker ILD (e.g., see Durlach and Colburn, 1978, p. 433).

In general, the maximum difference in IPD cues for T and M arises when the ITDs for T and M differ by one-half the period of the T frequency. For a 500-Hz T, the ITDs in T and M need to differ by roughly 1 ms to maximize binaural unmasking. For a 1000-Hz T, the ITDs in T and M need to differ by roughly 500 μs. This explains the dependence of maximal binaural unmasking on T and M separation and frequency: results in Figure 3-5 show that an angular separation of about 90˚ causes T and M ITDs to differ by roughly one ms (maximizing IPD differences in T and M for a 500-Hz T) whereas an angular separation of about 45˚ causes T and M ITDs to differ by roughly 500 μs.

3.5 Comparison of binaural unmasking to model predictions

3.5.1 Analysis

Subject-specific predictions of binaural unmasking were calculated using a modified version of the Colburn (1977a) model (a description of the current implementation of the model is provided in the Section 3.7 Appendix). Predictions depend on six parameters, evaluated at the T frequency: the ITDs and ILDs in both T and M; the binaural sensitivity of the listener; and the masker spectral level at the louder ear relative to absolute, monaural detection threshold, in quiet (parameter K).

The ITDs and ILDs used in the predictions were taken from the analysis of the cues present in the HRTFs. The ITD and ILD in M were calculated from the values averaged over the ERB filter centered on the T frequency (see Figure 3-5). The ITD and ILD in T were taken directly from the HRTF values at the T frequency (not averaged over the ERB). Binaural sensitivity at each frequency was set to the measured BMLD for each subject (Table 3-1). For both the 500- and 1000-Hz targets, the value of the K parameter was set to 44 dB/Hz.

3.5.2 Results

Model predictions are plotted alongside behavioral estimates of the binaural contribution to spatial unmasking in Figure 3-6 and Figure 3-7 (for the 500-Hz and 1000-Hz targets, respectively). In order to be somewhat conservative in identifying conditions where the model fails to account for behavioral data, a range of ±1 dB is shown around the actual model predictions.

Model predictions of binaural unmasking are non-negative for all spatial configurations. Predictions are exactly zero whenever T and M are at the same spatial location and positive whenever T and M have differences in either their IPDs or ILDs at the target frequency. Thus, in theory predictions of binaural unmasking are positive whenever T and M are at different distances but in the same direction (laterally displaced from the median plane) due to differences in ILDs in T and M. However, in practice, predictions are near zero for all configurations when T and M are in the same direction for Subjects S2, S3, and S4. Predictions for Subject S1 (who has very large ILDs for 15 cm sources and who had the largest BMLDs at both frequencies) are greater than zero for both T frequencies when T and M are at different distances but the same (off-median-plane) direction.

Binaural unmasking predictions are generally larger at 500 Hz than 1000 Hz. At both frequencies, binaural unmasking varies with angular separation of T and M; however, the angular separation that maximizes the predicted spatial unmasking depends on frequency. As in the behavioral results, predicted binaural unmasking is greatest when T and M are separated in azimuth by 90˚ for the 500-Hz T and 45˚ for the 1000-Hz T, corresponding to separations that maximize the differences in T and M IPD at the target frequency.

Also consistent with behavioral results, the maximum predicted amount of binaural unmasking decreases with masker ILD. As a result, the largest predicted amount

of binaural unmasking varies with M location (from panel to panel), systematically decreasing with M angle and decreasing when M is at 15 cm compared to 1 m.

Model predictions capture much of the variation in binaural unmasking; however, there are systematic prediction errors that are large compared to the intra-subject variability. (Note that the standard error in the behavioral results is less than or equal to 1 dB due to the experimental procedure. The error bars in the figure are more conservative, showing the *range* of thresholds obtained over multiple runs.)

Details of how predictions compare to behavioral results are first examined for the 500-Hz target (Figure 3-6). Predictions for Subject S1 agree well with behavioral results when M is at (0˚, 15 cm) and are fit reasonably well (with only one data point falling outside the model prediction range) for three other M locations [(45˚, 15 cm), (90˚, 15 cm), and (90˚, 1 m)]. However, S1 predictions tend to overestimate binaural unmasking for two M locations [(0˚, 1 m) and (45˚, 1 m)]. For Subject S2, predictions match behavioral results reasonably well when M is at 0˚, independent of M distance (although there are isolated data points for which the model overestimates binaural unmasking), but systematically underestimate binaural unmasking when M is at 45˚ and 90˚ (for both M distances). Results for Subject S3 are similar to those of Subject S2: predictions are in good agreement with measurements when M is in the median plane but underestimate binaural unmasking when M is laterally displaced.

Focusing on the 1000-Hz results (Figure 3-7), Subject S1 predictions generally overestimate binaural unmasking. For Subject S2, predictions generally underestimate binaural unmasking, except when M is at (45˚, 1 m), where predictions and measurements are reasonably close. Finally, predictions for Subject S4 either fit reasonably well or underestimate binaural unmasking when M is at 0˚ but overestimate binaural unmasking when M is at 45˚ or 90˚ (independent of M distance).

Overall, predictions and behavioral results are in better agreement when M is in the median plane than when M is at 45˚ or 90˚ and for the 500-Hz data compared to the 1000-Hz data.

3.5.3 Discussion

The Colburn model assumes that a single value representing binaural sensitivity at a particular frequency can account for inter-subject differences in binaural unmasking. This binaural sensitivity parameter was set from BMLD measures taken with a diotic M and target that was either diotic (NoSo) or inverted at one ear to produce an interaural phase difference of π (NoSπ). These conditions are most similar to the spatial configurations in which M is directly in front of the listener (and M is essentially diotic). For most of the configurations with M at 0˚, model predictions agree well with observed results. In contrast, larger discrepancies between the modeled and measured results arise when M is at 45˚ and 90˚ (conditions in which there are significant ILDs in M).

While there are some conditions in which the model predictions consistently over- or underestimate binaural unmasking (e.g., results for Subject S1 at 1000 Hz or for Subject S2 at 1000 Hz), there are other conditions for which changing the single subject-specific "binaural sensitivity" of the model cannot account for discrepancies between the model predictions and the measurements (e.g., results for Subject S2 at 500 Hz or for

Subject S4 at 1000 Hz). The current results suggest that subjects differ not only in their overall sensitivity to binaural differences, but also in the dependence of binaural sensitivity on the interaural parameters in M and/or T.

Even though there are specific conditions for which predictions fail to account for the results for a particular subject, the model captures many of the general patterns in results, including how the amount of binaural unmasking depends on the angular separation of T and M as well as the frequency of T and the tendency for binaural unmasking to decrease as the ILD in M increases.

3.6 Summary and conclusions

The current study is unique in measuring spatial unmasking when T and/or M are very close to the listener. Results show that for sources very close to the listener, small changes in source location can lead to large changes in detection threshold. These large changes arise due to large changes in both the TMR (affecting the energetic contribution to spatial unmasking) and ILDs (affecting the binaural contribution to spatial unmasking).

The current results demonstrate how the relative importance of energetic and binaural contributions to spatial unmasking change with T and M location, including source distance (in contrast to previous studies that considered only angular separation of relatively distant sources). For nearby sources, the relative importance of energetic contributions to spatial unmasking increases as M distance decreases, probably due to increases in the ILD in M, which reduce the amount of binaural unmasking. The energetic contribution also increases as T distance decreases, primarily because the TMR changes more rapidly with T angle when T is near the listener. The relative importance of the energetic contribution to spatial unmasking increases with T frequency, both because the absolute magnitude of energetic factors increases and because the binaural contribution to unmasking decreases. For a 500-Hz T, binaural and energetic factors are roughly equally important when M is in the median plane. However, energetic factors become relatively more important as M is displaced laterally, in part because the amount of spatial unmasking decreases with masker ILD. This trend, which is predicted by the Colburn model, helps to explain large differences in the amount of spatial unmasking observed in previous studies (e.g., Ebata et al., 1968; Gatehouse, 1987; Santon, 1987). Specifically, more spatial unmasking arises when M is positioned in front of the listener and T location is varied (leading to near-zero ILDs in M) than when T is fixed in location and the angle of M is varied (leading to progressively larger ILDs in M with spatial separation of T and M).

Binaural processing contributes as much as 10 dB to spatial unmasking for the spatial configurations tested (a value much larger than the 3 dB reported by Doll and Hanna, 1995). In theory, differences in T and M distance cause differences in T and M ILD when the sources are off the median plane, producing binaural unmasking. However, in the current study evidence of binaural unmasking due to differences in T and M distance were observed only for Subject S1, who had both the largest BMLDs and the largest ILDs of the four subjects in the study.

Although monaural detection thresholds were not directly measured in the current study, binaural performance is always better than or equal to the performance predicted by analysis of the TMR at the better ear. Thus, the current results do not help to explain

previous results suggesting that binaural performance sometimes falls below monaural performance using the better ear alone, particularly for configurations with large ILDs (Bronkhorst and Plomp, 1988; Shinn-Cunningham et al., 2001). One important distinction between the current study and these previous reports is that the current study measured tone detection for relatively low-frequency tones, whereas both of the previous studies measured speech intelligibility, a suprathreshold task that emphasizes information at higher frequencies. Further studies are necessary to help determine when binaural stimulation may actually degrade performance compared to monaural, better-ear performance.

Inter-subject differences in the amount of spatial unmasking are large and arise due to individual differences in 1) HRTFs, 2) overall binaural sensitivity, and 3) the way in which binaural sensitivity varies with spatial configuration of T and M. The Colburn model of binaural processing predicts overall trends in behavioral measures of binaural unmasking, but fails to capture subject-specific variations in performance. The spatial configurations for which model predictions are least reliable are the positions for which large ILDs arise in M and/or T, conditions that have not been extensively tested in previous studies. The current results suggest that the Colburn model must be modified so that subject differences in binaural sensitivity vary not only in overall magnitude but as a function of the interaural differences in M.

While the current model predictions cannot account for some small, but significant inter-subject differences, rough predictions of the amount of spatial unmasking capture most of the observed changes in detection threshold with spatial configuration. For instance, generic acoustic models of HRTFs (e.g., KEMAR measurements or spherical-head model predictions) combined with predictions of binaural unmasking using "average" model parameters should produce predictions that fall within the range of behavior observed across a population of subjects.

3.7 Appendix

A modified version of the model presented in Colburn (1977b) was used to predict the amount of binaural unmasking, defined as the difference in detection thresholds when T and M are at the same spatial location and when they are in different locations. The predicted amount of binaural unmasking is computed as

$$s(\alpha_T,\phi_T,\alpha_M,\phi_M,BMLD,K)=$$

$$\sqrt{\max\left(1,\frac{\alpha_T^4}{\alpha_M^4}\right)+\left(2*10^{\frac{BMLD}{10}}-1\right)R\left(\alpha_M,10^{\frac{K}{10}}\right)\frac{F^2(\phi_M,f_0)}{16}\left(1+\frac{\alpha_T^2}{\alpha_M^2}-2\frac{\alpha_T}{\alpha_M}\cos(\phi_M-\phi_T)\right)^2} \quad \text{(A1)}$$

where $\alpha_T=10^{ILD_T/20}$; $\alpha_M=10^{ILD_M/20}$; ILD_T and ILD_M are the interaural level differences in T and M (respectively) in dB; ϕ_T and ϕ_M are the IPDs of T and M (respectively) in radians; BMLD is the (subject-specific) binaural masking level difference in dB; K is the level of M relative to absolute detection threshold in quiet, in dB; and the functions F^2 and R are defined below (all evaluated at the T frequency).

Function F^2 represents the extent to which phase shifts in M cannot be compensated by internal time delays. This function is given by

$$F^2(\phi_M, f_0) = \frac{\sum_{k=-1000}^{1000} p\left(\frac{\phi_M}{2\pi f_0} + \frac{k}{f_0}, f_0\right) \exp\left\{-G^2(f_0)\left[1-\gamma\left(\frac{\phi_M}{2\pi f_0} + \frac{k}{f_0}\right)\right]\right\}}{\sum_{k=-1000}^{1000} p\left(\frac{k}{f_0}, f_0\right) \exp\left\{-G^2(f_0)\left[1-\gamma\left(\frac{k}{f_0}\right)\right]\right\}}. \quad \text{(A2)}$$

where p(τ, f) represents the relative number of interaural coincidence detectors (i.e., neurons in the medial superior olive) tuned to ITD τ and frequency f; G(f) represents the synchrony of firings of the auditory nerve at frequency f; and γ(τ) is the envelope of the autocorrelation function of the auditory nerve fiber impulse response at autocorrelation delay τ. In the current realization of the model, function p(τ,f) was modified to allow for a frequency-dependence in the distribution of interaural coincidence detectors (as suggested by Stern and Shear,1996), using

$$p(\tau, f_0) = \begin{cases} C\left(e^{-2\pi k_l|0.2|} - e^{-2\pi k_h|0.2|}\right)/|0.2| & |\tau| < 0.2ms \\ C\left(e^{-2\pi k_l|\tau|} - e^{-2\pi k_h|\tau|}\right)/|\tau| & |\tau| \geq 0.2ms \end{cases}$$

$$k_h = 3\,\mathrm{x}\,10^6;\; k_l = \begin{cases} 0.1\left(f_0\ 10^{-3}\right)^{1.1} & f_0 \leq 1200\ Hz \\ 0.1\left(1200\,\mathrm{x}\,10^{-3}\right)^{1.1} & f_0 > 1200\ Hz \end{cases} \quad \text{(A3)}$$

$$C = \begin{cases} 0.15336335 & f_0 = 500\ Hz \\ 0.19994680 & f_0 = 1000\ Hz \end{cases}.$$

G(f) is given by

$$G(f_0) = \begin{cases} \sqrt{10} & f_0 \leq 800\ Hz \\ \sqrt{10}\,\frac{800}{f_0} & f_0 > 800\ Hz \end{cases}. \quad \text{(A4)}$$

γ(τ) is given by

$$\gamma(\tau) = \begin{cases} 2.359\,\mathrm{x}\,10^{-4}\tau^5 + 1.5207\,\mathrm{x}\,10^{8}\tau^4 - 1.764\,\mathrm{x}\,10^{4}\tau^2 + 0.993 & |\tau| \leq 0.006 \\ -97.3236|\tau| + 1.139 & |\tau| > 0.006 \end{cases} \quad \text{(A5)}$$

where τ is in milliseconds.

Finally, function R(α, K) characterizes the decrease in the number of activated auditory nerve fibers in the ear receiving the less intense signal as a function of masker ILD. The current implementation uses a modified version of Eq. 35 from Colburn (1977b):

$$R(\alpha_n) = \begin{cases} \left(\frac{10\log_{10}\alpha_n^{-2}K}{40}\right)^2 & \alpha_n^{-2}K \leq 10^4 \\ 1 & \alpha_n^{-2}K > 10^4 \end{cases}. \quad \text{(A6)}$$

where K is the ratio of the spectrum level at the louder ear to the detection threshold in quiet. This implementation of the model assumes that the auditory nerve fibers at each T frequency have thresholds uniformly distributed (on a dB scale) over a 40-dB range above the absolute detection threshold for that frequency.

Chapter 4 Localization in reverberant rooms: effect of nearby walls and experience

Abstract

Localization accuracy was measured for sources in the right frontal quadrant of the listener's horizontal plane, at distances between 15 and 100 cm. Listeners were positioned at four different locations in the room: in the center, with their back close to a wall, with their left ear close to a wall, or in the corner. The mean (i.e., bias) and variance of the error in perceived azimuth and distance were analyzed. Results were influenced by experience and room position. Generally, variance decreased over time and increased with the acoustic complexity of the room position. Room position had no effect on perceived distance bias and a small but significant effect on the azimuthal bias. In particular, when the listener had his back to the wall, responses were biased medially. Experience affected the bias in both azimuth and distance for sources ahead of the listener, but not near the interaural axis. In addition, two of three listeners who started in the corner of the room showed a consistent distance bias, underestimating source distance. The effect of room position was stronger for nearby sources while the effect of experience was greater for far sources. No learning was observable within a single session. These results show that both experience and room positions affect auditory localization accuracy and reliability.

4.1 Introduction

This study evaluates the ability of human listeners to localize nearby sound sources in an ordinary reverberant classroom. Two factors are specifically examined: how listener's location in the room influences perception, and how the amount of experience with the task and room influences perception.

4.2 Background

Sound localization has traditionally been studied in an anechoic chamber, and has usually focused only on two of the three spatial dimensions: source azimuth and elevation (Wightman and Kistler, 1989; Makous and Middlebrooks, 1990; Wenzel et al., 1993). There are only a few studies that examined performance in reverberant environments, in either azimuth (e.g., Hartmann, 1983; Rakerd and Hartmann, 1985, 1986; Wagenaars, 1990) or, more recently, in distance (Bronkhorst and Houtgast, 1999; Santarelli, 2000; Zahorik, 2000). These studies show that reverberation provides distance information (Bronkhorst and Houtgast, 1999; Santarelli, 2000; Zahorik, 2000) but slightly degrades azimuthal localization accuracy (Santarelli, 2000), and that the degradation of azimuthal perception can be overcome as a listener gains experience with the room (Shinn-Cunningham, 2000).

Brown (2001) performed an analysis of the effects of the room reverberation on acoustic characteristics of the sounds reaching the listeners' ears. Using a KEMAR manikin, Brown measured the head-related impulse responses (HRIR) for several positions of the listener in a room ("room positions") and various positions of the sound source relative to the listener ("source positions"). Brown showed that all the important localization cues (monaural spectrum, interaural level difference or ILD, and interaural time difference or ITD) are influenced by reverberation. Brown also showed that the effect of reverberation depends on both room position and source position.

The anechoic acoustic cues for azimuthal perception are relatively well understood: the two most robust cues are the ITD and ILD, both of which grow with source laterality. Brown showed that both of these cues are altered by room reverberation. Specifically, reverberation increases the variance in both ITD and ILD across frequency. In addition, the ILD associated with a given source position can differ from one room position to another, for example, when there is a wall near the contralateral ear, the ILD is smaller. These results suggest than room reverberation may degrade localization accuracy in azimuth and increase variance in azimuth responses by introducing inconsistency in the internal cues over frequency. Similarly, the change in mean ILD with room position may cause a bias in perceived azimuth. However, Brown's analysis examined the effects of reverberation by looking at HRTF – which ignores how cues change over time. As a result, wile this analysis gives some insight into the effects of reverberation, it does not fully explain how perception is influenced by room acoustics.

Distance perception is hypothesized to be based on some correlate of the direct-to-reverberant energy ratio (D/R) in the perceived sound (Bronkhorst and Houtgast, 1999). For a given source distance, D/R changes with both room position and source angle relative to the listener. Thus, unless a listener uses a room-position and source-angle dependent mapping of D/R to distance, perceived distance might be biased when

the listener is at certain room positions. Moreover, for nearby lateral sources the ILD changes with source distance providing a potential cue for distance perception (Brungart, 1998; Shinn-Cunningham, Santarelli and Kopčo, 2000). Given that reverberation affects the ILD, distance perception should be affected by room acoustics if listeners base their distance judgments on ILD.

Santarelli, Kopčo and Shinn-Cunningham (1999a) performed a localization study in a real room and found that the listeners' performance improved over the course of the experiment, although no improvement was observed in a similar study performed in anechoic space (Brungart and Durlach, 1999). Because the rate of this learning was fairly slow (learning occurred over the course of days, across multiple sessions) Santarelli et al. (1999a) hypothesized that this improvement is due to the listeners' subconscious learning of the acoustic properties of the room. The current experiment was performed in the same room and over a similar time scale as the study by Santarelli et al. (1999a). Thus, a similar "room learning" effect was expected in the current study.

4.3 Experiment

The current study investigates the influence of room position (acoustic effects that vary with the position of the listener in the room) and room learning (changes in performance due to listener's experience in a particular room) on auditory localization. While seated at one of four positions in the room, listeners indicated the perceived source location of sounds originating in the horizontal plane at the level of their ears. Listeners were divided into two groups differing in the order in which the room positions were presented. Source location varied in both azimuth and distance relative to the listeners. Both the mean and standard deviation in the perceived azimuth and distance of the sources was evaluated to judge how experience and room acoustics influence localization.

4.4 Methods

4.4.1 Listeners

Six paid graduate students (three male and three female) participated in the study. Their ages ranged from 23 – 28 years. One subject had prior experience in auditory localization experiments. All six subjects had normal hearing as determined by an audiometric screening.

4.4.2 Stimuli and apparatus

Stimuli consisted of five 150-ms-long pink-noise bursts separated by 30-ms-long gaps of silence. One of five random tokens of the stimulus was chosen for presentation in each trial. The stimuli had a wideband pink frequency spectrum (roll-off 6 dB/octave from 200-15kHz) and a 120-dB/octave roll-off out of band.

A point source (Brungart, 1998) was used to present the stimuli, which were corrected for the non-flat spectral response of the point source by convolution with a linear-phase filter with a spectrum equal to the inverse of the point source's response. To eliminate overall loudness as a distance cue, the RMS level of the stimulus was crudely normalized on each presentation by attenuating it proportionally to the distance from the head, then roving it by an additional ±7.5 dB. With the rove, the stimulus level at the

nearer ear was random and uniformly distributed between 44 and 59 dBA from trial to trial.

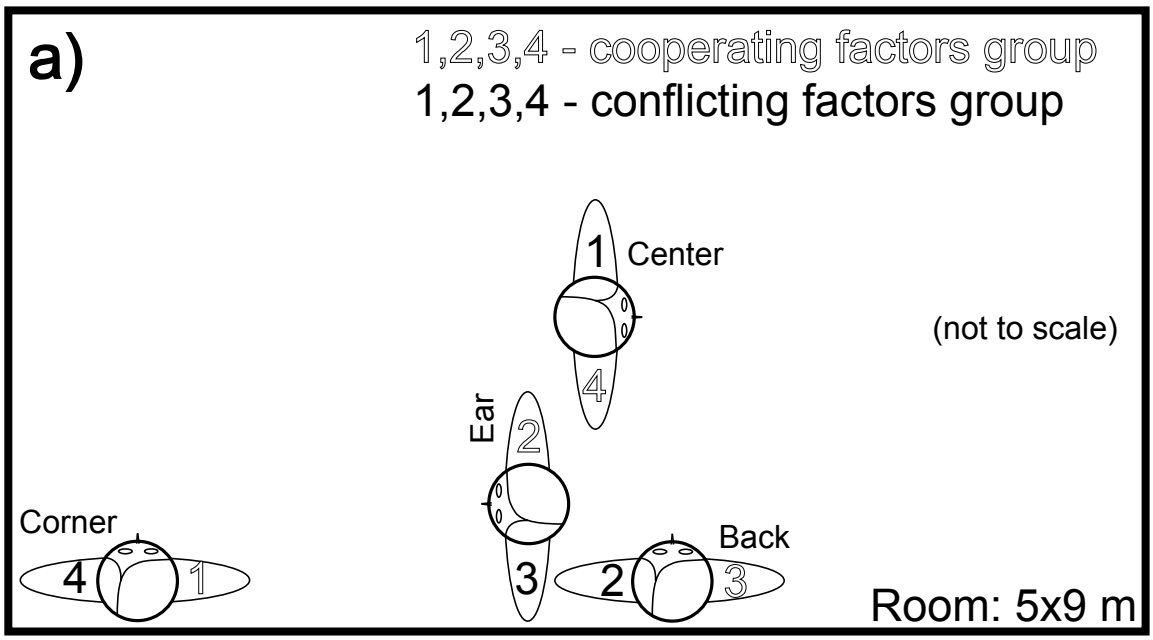
Five pre-generated stimulus files were stored on the hard disk of the control computer. On each trial, one of the samples was randomly chosen, appropriately re-scaled, and sent to the point source via a Cirrus CS4236 16-bit stereo sound card and a Crown power amplifier. A Polhemus FastTrak electromagnetic tracker was used to record the position of the subject's head, the sound source position, and the subject's response in each trial.

4.4.3 Procedure

The experiment was performed in an ordinary, empty classroom (size 5 x 9 meters, T_{60} = 0.7 s, Figure 4-1a). Each subject performed four two-hour sessions, each from a different location in the room. These positions, shown in Figure 4-1a, differed in how close the nearest walls were relative to the subject. In the "center" location, the subject was far from any walls roughly in the middle of the room. In the "back" location, the subject was seated with his/her back facing the longer of the walls approximately in the middle of the wall (distance from the subject's back to the wall was approximately 15 cm). In the "ear" location, the subject was in the same location as in the "back" location, except that the subject was rotated by 90° so that his/her left shoulder was approximately 15 cm from the wall. Finally in the "corner" location the subject was seated in the corner of the room with his/her shoulder approximately 15 cm from the shorter wall and his/her back approximately 15 cm from the longer wall.

The subjects were divided into two groups depending on the order in which the room positions were tested. The *conflicting factors* group started in the center and ended in the corner positions (i.e., the subjects started in the acoustically simplest condition and ended in the most acoustically complex condition). For this group, the room learning and room position factors were expected to be in conflict. Specifically, room learning should lead to better performance in the last session, which was in the corner, however, the acoustic cues should lead to better performance in the center, which was performed first. The second, *cooperating factors* group started in the corner and ended in the center positions. For this group, the effects of both room learning and room position were expected to lead to better performance from session to session.

In each of the four sessions, 300 trials were performed separated by breaks every 50 trials. A practice session of 50 trials preceded the first session. Each trial started by the subject closing his/her eyes, after which the experimenter placed the source at a random, computer-generated position. The stimulus was then presented after which the experimenter moved the source to a neutral position. The subject was then allowed to open his/her eyes and point to the perceived location of the sound using a wand on which an electromagnetic position sensor was mounted. The sound source positions were distributed uniformly in azimuth in one of three 15°-wide bins (around 0, 45, or 90°) shown in Figure 4-1b. The distance dimension was logarithmically distributed.



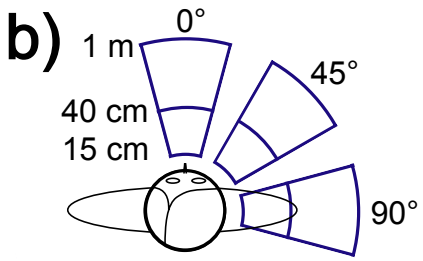


Figure 4-1 a) Listener positions in room. The order of positions for the two subject groups (cooperating factors group and conflicting factors group) is shown by numerals and distinguished by font type (normal vs. outlines). b) Bins of locations for source presentation.

4.5 Results and discussion

Results are analyzed in the dimensions of azimuth and distance. The azimuth error is defined as a (signed) difference between actual and perceived azimuth on a given trial. The distance error is defined as $\log_{10}$(perceived distance / actual distance) on a given trial. Both the mean and standard deviations in the azimuthal and distance errors are analyzed and discussed.

4.5.1 Initial vs. final session

4.5.1.1 Results

Both factors of interest in this study, room position and room learning, have the largest effect on localization when comparing the initial vs. the final session. The effect of room learning is maximized between these two sessions because the sessions are maximally separated in time. Room position effects are maximized because the initial and final sessions were performed in the center vs. the corner of the room, that is, in the acoustically simplest vs. the most complex conditions. Figure 4-2 shows the results (mean and standard deviation in perceived azimuth and distance) in the initial vs. the final session. The thick lines plot the across-subject average performance for each subject group; the thin lines plot individual subject data. The conflicting factors group did the initial session in the center and the final session in the corner position. For the cooperating factors group the room position the ordering was reversed.

The computation of the statistics shown in Figure 4-2 was as follows. For each subject and each room position, the error in each trial was computed as the (signed) difference between the perceived and actual azimuth (or log distance). Then the mean (i.e., bias) and standard deviation in this error was computed within each of the six source bins from Figure 4-1b (separately for each subject and each room position). Then, for each subject, the mean of these values (of the previous mean and standard deviation) was computed across the six source bins and these values are plotted as individual subject data in Figure 4-2. The group averages were computed by taking the mean of the corresponding values for the three subjects in a given group.

Panel a in Figure 4-2 shows how the mean azimuthal bias changes between the two sessions. Overall, there is no consistent change in the bias in either subject group. Two of the three conflicting-factors subjects show a decrease in bias from the initial to the final session; two of the three cooperating-factors subjects show an increase from the initial to the final session. The across-subject average bias change from the initial to the final session is small in both groups relative to the intersubject variability.

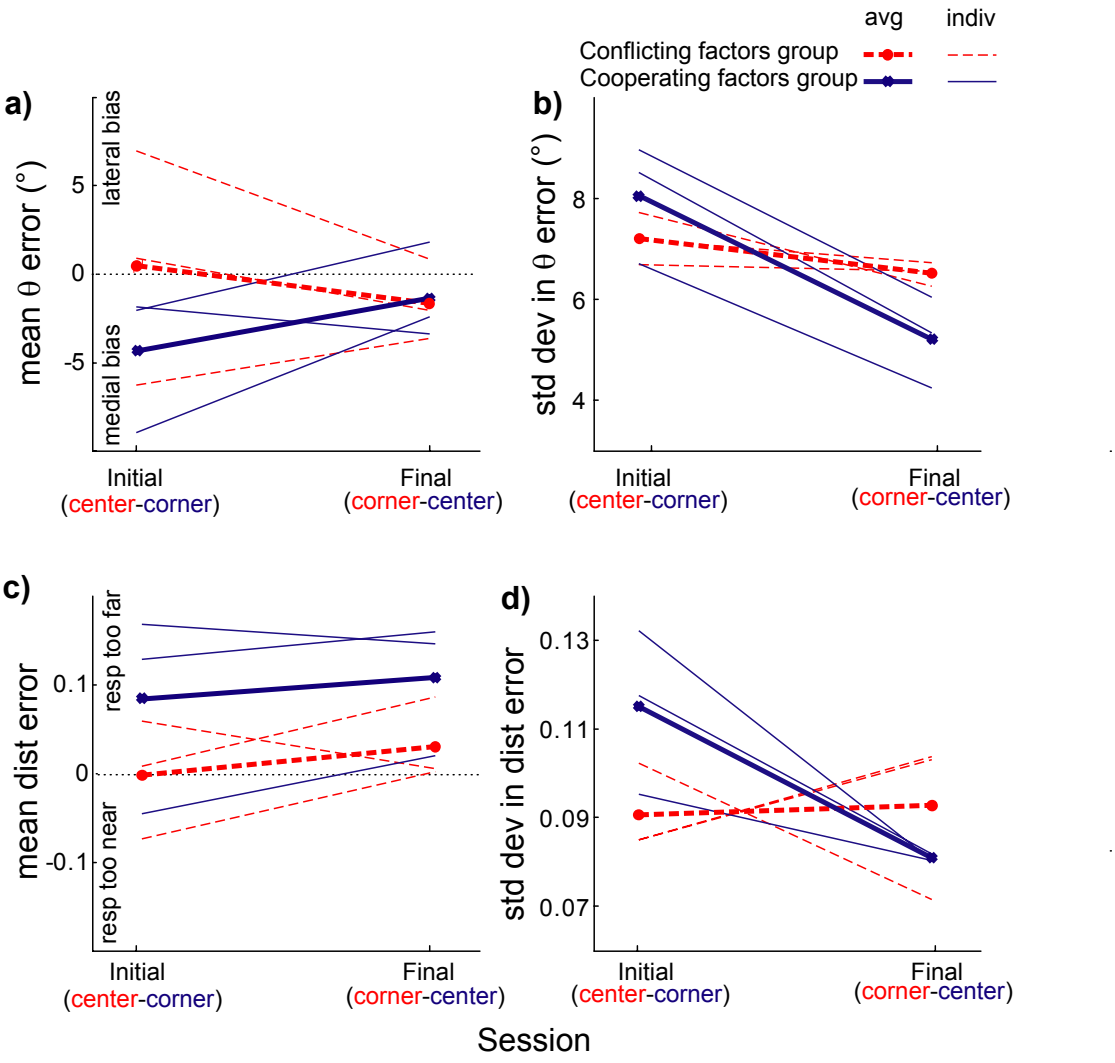


Figure 4-2 Average bias and variability in signed response errors (averaged over source position) in the initial and the final experimental session. The conflicting factors group started in the center and ended in the corner position. This order was reversed for the cooperating factors group. a) mean azimuth bias computed as perceived - actual azimuth, in degrees; b) standard deviation in azimuth response; c) distance response bias computed as $\log_{10}$(perceived dist / actual dist); d) standard deviation in distance response.

Figure 4-2c shows the average bias in perceived distance. The largest effect is an overall difference between the two subject groups: on average, the conflicting factors group overestimated distance by 3.6%, whereas the cooperating factors group overestimated distance by 20.2%. This effect is caused by a strong and consistent tendency for two of the subjects in the cooperating factors group to overestimate distance; independent of the session order or room position (the third cooperating-factors subject had a smaller bias). In general, there are large intersubject differences in distance perception (see Zahorik, 2002). Thus, the most parsimonious explanation for this difference between groups is that it is due to subject differences. There are no other clear trends in the overall distance bias.

The standard deviation in azimuth and distance responses is shown in Figure 4-2 panels b and d, respectively. Results show that there is an interaction between room learning and room position. For the cooperating factors group (full lines) the variability always decreases over time, whereas for the conflicting factors group (dashed lines) the variability is essentially constant relative to intersubject differences. These results suggest that room learning reduces response variability over time; however, variability in responses increases with the acoustic complexity of the signals reaching the listener (i.e., with room position).

4.5.1.2 Discussion

The rate of the learning effect observed in the standard deviation measures is fairly slow, similar to the results of Santarelli (2000). This confirms that whatever the listener learns while in one room position generalizes to other positions in the room.

The effect of the change in the room position on response variability can be explained by the acoustical properties of the perceived sound. In the center of the room there are no nearby walls and the received reverberation is essentially independent of the sound source position and relatively less intense than for other room positions. In the corner, the pattern of reverberation is dominated by early reflections from nearby walls, the magnitude of which can sometimes be comparable to the magnitude of the direct sound. Since the intensity and timing of the early reflections change dramatically as a function of source position, it is not surprising that such reflections will increase response variability.

4.5.2 Room position vs. room learning: detailed results

Figure 4-3 presents a detailed analysis of the results of this behavioral study. The computation of the statistics shown in the figure was as follows. For each subject and each room position, the error in each trial was computed as the (signed) difference between the perceived and actual azimuth (or log distance). Then the mean (i.e., bias) and standard deviation in this error was computed within each of the six source bins from Figure 4-1b (separately for each subject and each room position). These values (the mean and standard deviation) were averaged across the subjects in each group to get the group means in columns 2 to 7 (thin lines). For each room position, averages across the values in columns 2-7 were computed and plotted in the corresponding panel 1. In addition, averages across the groups were computed and plotted in each panel (thick lines). Any

differences between the subject groups can be attributed to either subject differences or an effect of room learning.

4.5.2.1 Mean azimuthal error

Figure 4-3 panel a shows the mean difference between the actual and perceived azimuth of the sound sources. Averaged across the source bins (panel a1) there is no effect of learning (full vs. dashed thin lines). Data analyzed separately by the source bin (panels a2-a7) show interaction between the source azimuth and source distance. Averaged across the room positions, there is no bias for near lateral sources (panel a4) and negative bias for the other near bins (panels a2-a3). For far sources, there is no bias for the medial (panel a5), negative bias for the intermediate (panel a6), and positive to zero bias for the lateral sources (panel a7).

The average graph (panel a1) shows also a small effect of room position on the bias: the back and corner positions are biased negatively while the center and ear positions have no bias. Also, there is an interaction between the room learning (thin lines) and the room position for medial sources (panels a2 and a5) exhibited in the relative changes in bias between the two groups in the back and ear positions. However, these effects are small compared to the within-subject and inter-subject variance in the data.

Figure 4-3 Localization performance as a function of the listener position in the room (Center, Back, Ear, and Corner). The first column shows the average values across subjects and source position. Columns 2-7 show across-subject averages for different source position bins (shown in Figure 4-1). a) mean azimuth bias computed as perceived - actual azimuth, in degrees; b) standard deviation in azimuth response; c) distance response bias computed as $\log_{10}$(perceived dist / actual dist); d) standard deviation in distance response.

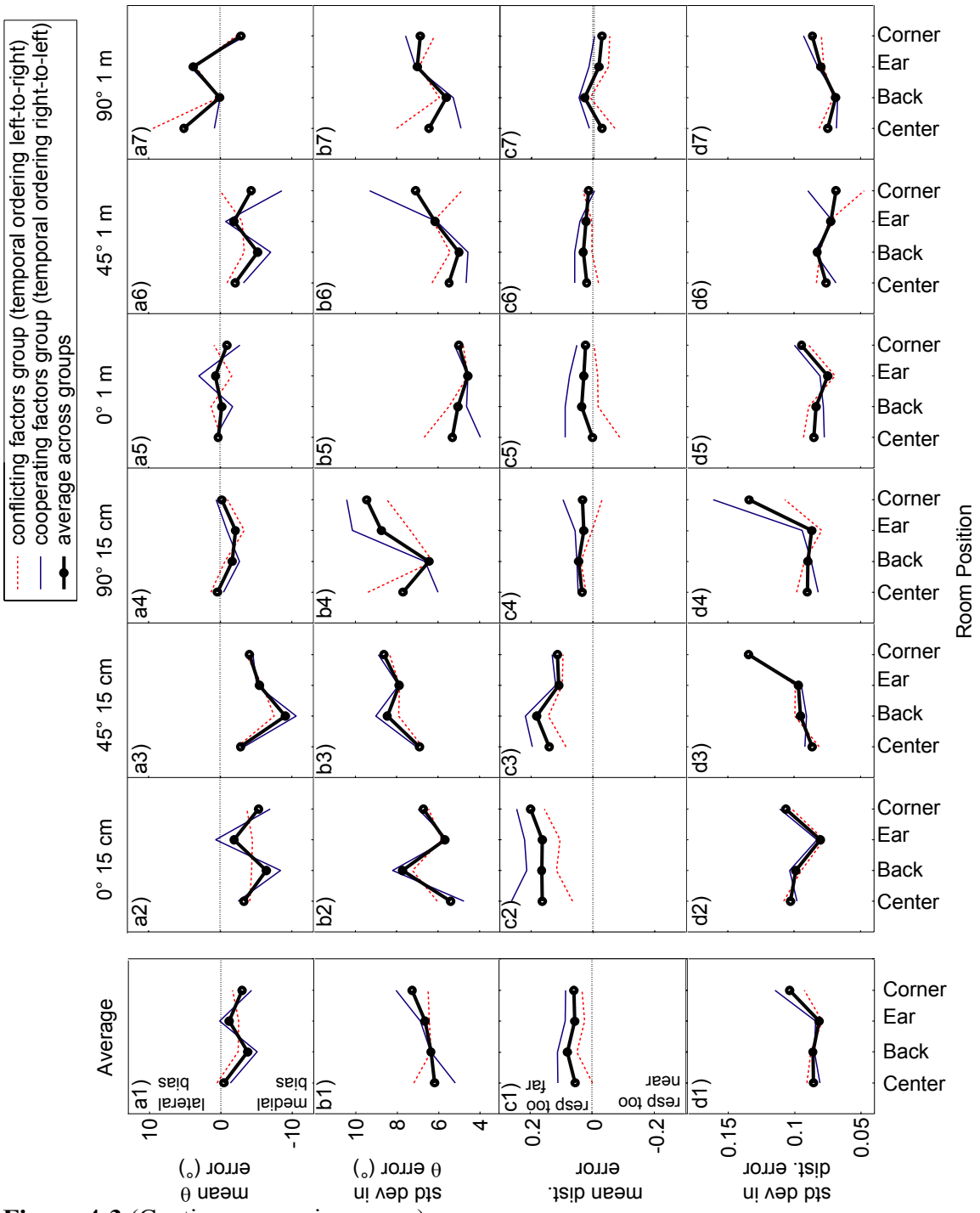


Figure 4-3 (Caption on previous page)

4.5.2.2 Variance in azimuthal perception

Graphs in Figure 4-3b show the standard deviation in azimuthal error. Panels b2-b7 show the standard deviation separately for each source position bin. These panels show that standard deviation grows with source azimuth (panel 2 vs. 3 vs. 4 and 5 vs. 6 vs. 7) and is larger for near sources (panels 2 to 4) than for the far sources (panels 5 to 7). In some regions (far sources and near lateral sources, panels 4 to 7) there is large difference between the two subject groups, suggesting a possible effect of room learning.

Results are averaged across source position bins in panel 1. The across-group average (thick line) shows that the amount of variability in subjects' responses is largest for the corner and smallest in the center positions. This result can be explained by the presence of a number of early strong reflections that influence the perceived sound. For the corner position, the intensity and timing of the reflections changes dramatically with small changes in the sound source location. Thus, in addition to causing larger distortion of acoustic response variability cues, the early reflections alter cues in ways that depend on source and listener position, making this distortion difficult to overcome by learning. The gradual growth in the amount of response variability for the intermediate (back and ear) positions is consistent with this hypothesis. Comparing the two subject groups (thin lines) shows that room learning influences azimuth response variability. In particular, the conflicting factors group actually improves when moved from the center to the corner of the room, despite the increase in the complexity of the acoustic environment. This result confirms the conclusion from the overall analysis (Section 4.5.1) that the influence of experience on the variance in perceived azimuth is as strong as the influence of the room position.

4.5.2.3 Mean distance error

Graphs in Figure 4-3c plot the mean error in perceived distance. The average graph (Figure 4-3 panel c1) shows that the perceived distance error is essentially independent of the position of the listener in the room (consistent with results in Section 4.5.1). In addition, while there is a consistent difference between the cooperating- and conflicting-factor groups, this difference is independent of both experience and room location and thus is most likely due to subject differences. This result does not support most current theories of distance perception (Bronkhorst and Houtgast, 1999; Santarelli, 2000), which assume that the computation of perceived distance involves estimation of the amount of the direct-to-reverberant energy ratio. The amount of reverberant energy changes dramatically with the listener's position and, in some cases, with the source azimuth relative to the listener (Brown, 2001). However, neither of these factors influences distance error. These results suggest either that the actual acoustic cue that is perceptually important, while varying with reverberation level is independent of the room-position-related changes in the reverberation pattern, or that the brain recalibrates how distance is computed by the "reverberation cue" depending on distance and source positions.

Averaged across groups, an effect of the source location can be observed (panels c2 to c7). Panels c2-c5 show that the overall bias to overestimate distance is actually present only for nearby sources (although for very distant sources, subjects tend to

underestimate distance, Zahorik, 2000) and that it decreases with azimuth. The fact that the bias changes with azimuth suggests that listeners cannot compensate for the changes in reverberation cue due to the source azimuth.

As was the case for the overall plot (panel c1), the results as a function of source location (panels c2-c7) averaged across subject groups (thick lines) show essentially no effect of room position. Comparison of the subject groups shows that the difference in bias between the groups is apparent only in some room and source positions (panels c2, c3, and c5). Also, the effect decreases with source laterality and with source distance.

4.5.2.4 Variance in distance perception

Figure 4-3d shows the standard deviation in distance response error. Averaged across source positions (panel 1), the standard deviation in distance error is largest when the listener is in the corner, and is smaller and essentially equal for the other three room positions. This result suggests that the variability in distance error is only affected by very large perturbations in the reverberation pattern.

When the data are binned by the sound source position (Figure 4-3 panels d2-d7), no effect of room position is visible for far sources (panels d5-d7). The increased variance in distance perception for the corner position is driven only by an effect for nearby, lateral sources (panels d3-d4). This effect could be partially explained by the inaccuracy in the listener's pointing to perceived position, which is larger for near than for far sources, however, the fact that the error is independent of distance for three out of the four room positions suggests that the increased variance in the corner of the room is an effect caused by the room acoustics. There are only two points where there is a large difference in performance between the two subject groups (panels d4 and d6), which supports the conclusion that the effect of room learning on the standard deviation in distance error is smaller than the effect on standard deviation in azimuth error.

4.5.3 Effect of room learning for near vs. far sources

Figure 4-4 presents an analysis of the interaction between the effect of learning and the listener's position in the room as a function of the sound source distance. The change in the mean and standard deviation in the azimuth and distance errors are shown. In each panel, the data for the two different subject groups are shown as histograms. The histograms were generated as follows. For each subject and for the center and corner room positions, the error in each trial was computed as the (signed) difference between the perceived and actual azimuth (or log distance). Then the mean (i.e., bias) and standard deviation in this error was computed within each of the six source bins from Figure 4-1b (separately for each subject and each room position) and the difference between these values (the mean and standard deviation) was computed between the initial and the final session (i.e., corner - center for the conflicting-factors group and center - corner for the cooperating-factors group). Histograms of these differences were generated. Three separate histogram pairs are shown in each panel, one for the overall performance (left panel), one for near (center panel) and one for far (right panel) sources. The left-most graphs in each panel show the distribution of the 36 points (6 subjects x 6 source position

bins) that were averaged to generate the plots in Figure 4-2. The center and right-most graphs show the distribution of 18 points (6 subjects x 3 source position bins) separately for near or far sources.

Results show that there is no clear change in bias either for azimuth or distance (the distributions are centered on zero), and no difference between subject groups (panels a and c). However, response variability tends to decrease over time for booth the cooperating-factors group and the conflicting-factors group (panels b and d). This effect is larger for the cooperating-factors group than the conflicting-factors group.

Comparing near and far sources shows that for far sources, the decrease in response variability is essentially equally large for both groups. For near sources the decrease is still present in the cooperating-factors group data, while the data are centered at zero for the conflicting-factors group. This suggests that the effect of room learning increases with source distance from the listener.

Figure 4-4 Effect of source distance on learning. Histograms show the change in performance between the initial and the final session for the conflicting factors group (dashed lines) and the cooperating factors group (full lines). Separate histograms in each panel show overall performance (left-hand graph), performance for near sources (center graph), and far sources (right-hand graph). Changes are computed as differences in a given parameter between the initial and the final session. a) change in the mean azimuth biases (mean azimuth bias computed as perceived - actual azimuth, in degrees); b) change in standard deviation in azimuth response; c) change in distance response bias (distance response bias computed as $\log_{10}$(perceived dist / actual dist)); d) change in standard deviation in diastase response.

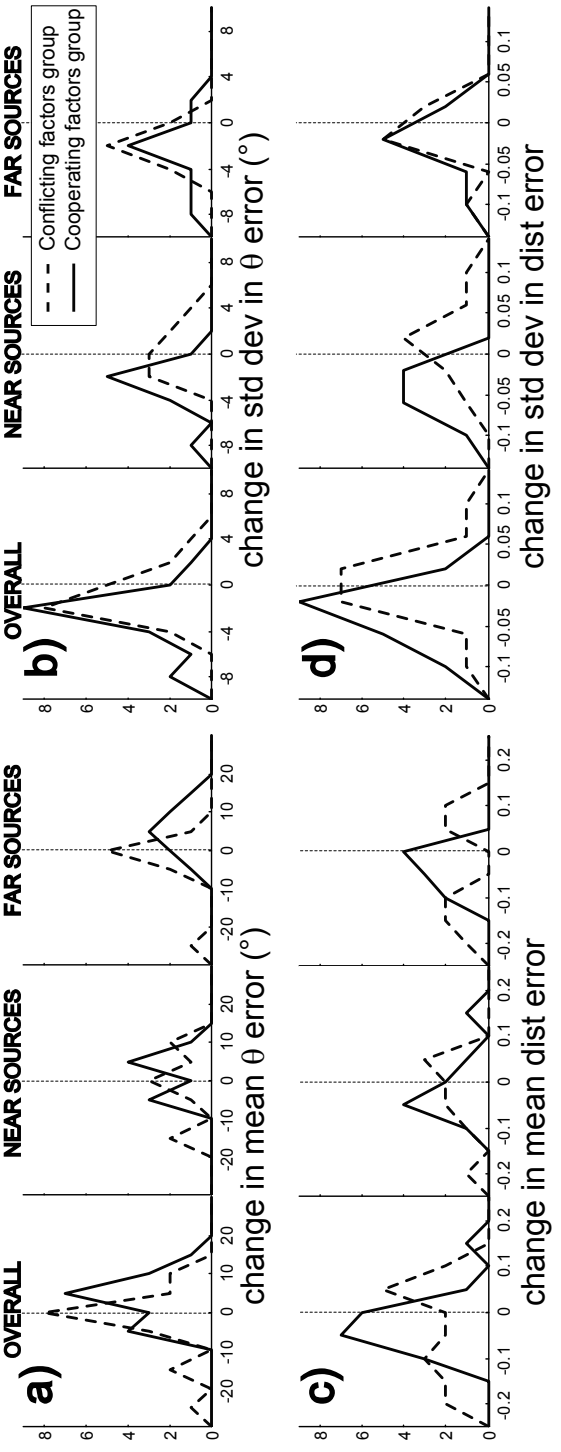


Figure 4-4 (Caption on previous page)

4.5.4 Learning within a session

Preceding analysis shows that there is a room learning effect when comparing the initial vs. the final experimental session. The present analysis evaluates whether the learning is visible on a much smaller scale, corresponding to one continuous session, i.e., 1-2 hours. Figure 4-5 plots for each subject the difference in the mean and standard deviation of perceived azimuth and difference between the first and the second half of each experimental session. No systematic trends are observed. This result suggests that room learning is apparent only for time periods larger than two hours, or that breaks are required between sessions in order to observe any learning effect.

4.6 Summary and conclusions

The results of this study show that both the position of the listener in the room and the listener's experience with the room influence the listener's ability to accurately localize sounds. The main effect of both these factors is to affect the variability in responses. Decease in response standard deviation is observed over time, presumably due to the room learning effect. Also, the standard deviation is larger in the corner of the room than in the center, suggesting that when the room reverberation is more complex (e.g., in the corner) the variance in perceived location is larger. While the room location is roughly equally influential on variability in azimuth and distance, the room learning has a stronger effect on azimuth than distance.

Contrary to expectations based on current theories of distance perception, there is no effect of room position on the perceived distance bias. On the other hand, room learning influences distance bias for lateral sources.

Finally, the room learning effect, while clear across several sessions (on a time scale of 6-8 hours performed over a course of weeks), is not observed within a session (on a time scale of 1-2 hours).

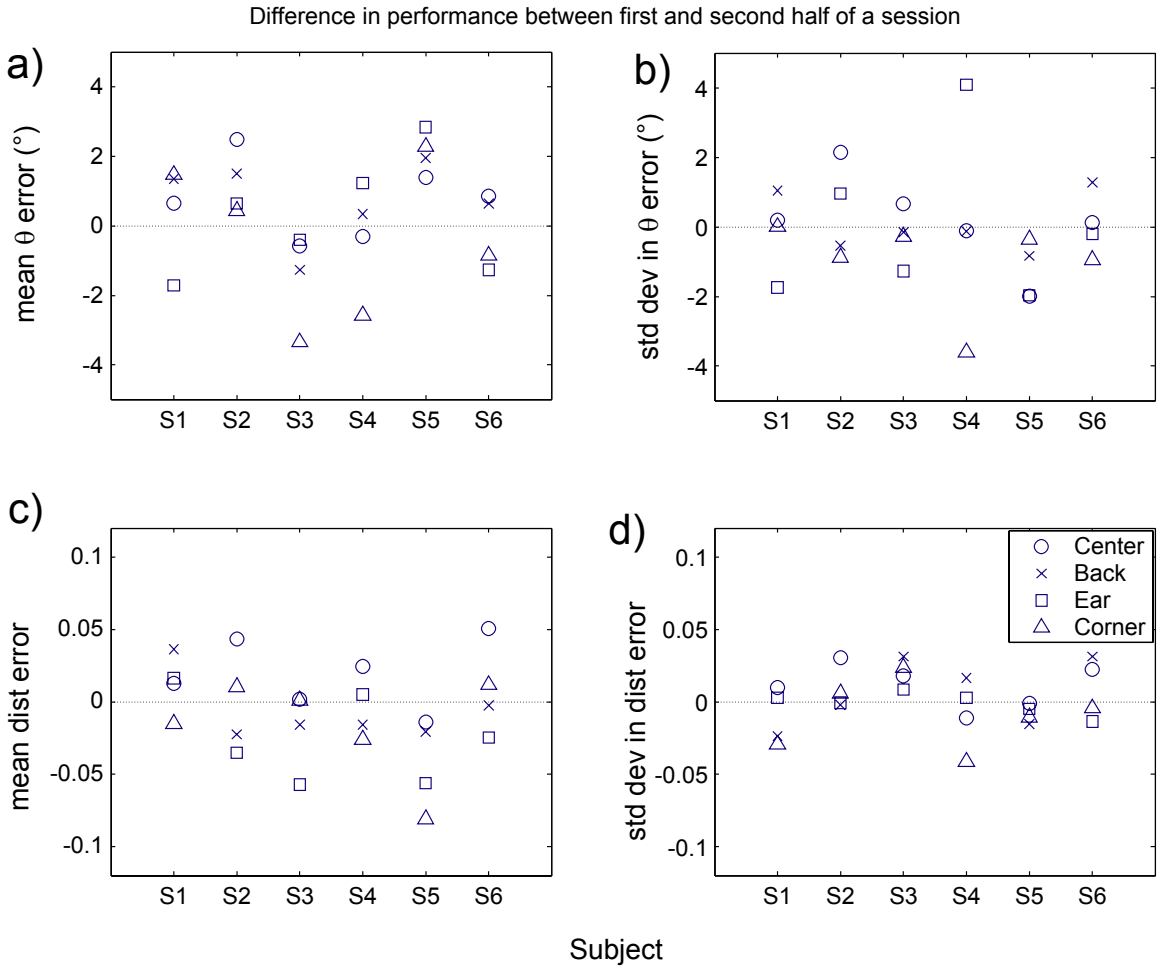


Figure 4-5 Effect of learning within a session. Difference in the mean and standard deviation in perceived azimuth and distance between the first vs. the second half of the experimental session.

Chapter 5 Auditory Localization in Rooms: Acoustic Analysis and Behavior[1]

Abstract

In an ordinary room, reverberation and echoes in the signals reaching a listener's ears influence auditory localization performance. The energy of the echoes and reverberation depends on the position of the listener in the room as well as on the position of the sound source relative to the listener. In this chapter, the effects of echoes and reverberation are quantified through analysis of reverberant Head-Related Transfer Functions (HRTFs) measured in an ordinary classroom. HRTFs were measured for several human listeners and a KEMAR acoustic manikin at four different listener positions in the room and multiple source positions relative to the listener. Azimuthal localization performance was also measured for several listeners in the room as a function of listener position. Compared to the acoustic cues this performance was found to be less sensitive to a change in room location. The only similarity was found between the magnitude of frequency-to-frequency variations in basic localization cues and the variability in localization performance, demonstrating that localization accuracy decreases with increasing reverberant energy.

5.1 Introduction

In a room, the ability of human listeners to localize sounds is influenced by echoes and reverberation (which are henceforth collectively referred to as "reverberation," for brevity; Santarelli, 2000). The effect of reverberation can be both beneficial and detrimental, improving distance perception and degrading azimuthal localization. However, the pattern of reverberation differs from room to room as well as from position to position within a given room. For a listener in the center of a room, most reflective surfaces are relatively far from the listener and reflections are diffuse for all source positions. On the other hand, when the listener is close to a wall, prominent early reflections arise whose magnitude and timing depend on the location of the source relative to the walls and to the listener.

The goal of this study is to analyze how localization cues in the signals reaching a listener's ears are influenced by reverberation and to evaluate whether acoustic effects can account for how localization performance varies with a listener's position in a room. A set of head-related transfer functions (HRTFs; see Santarelli, 2000) was measured for a manikin (KEMAR) located at different positions in a classroom. The effect of reverberation on interaural differences and spectral magnitude is evaluated by computing how these cues vary with source position relative to the listener and listener location relative to the room. Results are compared to behavioral localization results (Chapter 4, also Kopčo, Brown, and Shinn-Cunningham, 2001) for similar configurations of source and listener in the room.

---

[1] Published in the Proceedings of the 32nd International Acoustical Conference - EAA symposium "ACOUSTICS BANSKA STIAVNICA 2002" September 10 - 12, 2002

5.2 Methods

5.2.1 Acoustic analysis

HRTFs were measured for a KEMAR manikin located at the four positions in a classroom (Center, Back, Ear, Corner) (T60≈700 ms) shown in Figure 4-1.

HRTFs were measured for sources in KEMAR's right front quadrant at all combinations of azimuths from 0° to 90° (15° steps) and distances 0.15, 0.40, and 1 m (for sources in the horizontal plane containing the ears). Responses to Maximum-Length Sequences (e.g., see Zahorik, 2000) were measured to estimate a 750-ms-long head-related impulse response (HRIR; 44.1 kHz sampling rate). Stimuli were presented from a PC computer using a TDT system, Crown amplifier, and a Bose cube speaker. Knowles Electret microphones mounted in earplugs in KEMAR's ear canals were fed back to the TDT to make blocked-meatus recordings. The magnitude spectrum of the measurement system was relatively flat (within 10 dB) between 300 Hz – 12 kHz range. The dynamic range was at least 40 dB at all the frequencies. HRTFs from the center-room position were time-windowed using a cosine-squared onset/offset window (1 ms) to obtain pseudo-anechoic HRTFs against which other measurements are compared.

Interaural level differences (ILDs) were computed as the difference between the left and right ear HRTF RMS energy between 2000 – 5000 Hz. ILD variability was computed as the mean absolute value of the frequency-to-frequency difference in the ILD (using a frequency step of 1 Hz). Interaural time differences (ITDs) were estimated from the interaural delay producing the maximum peak in the cross-correlation of the left- and right-ear HRIRs bandpass-filtered from 200 – 2000 Hz.

5.2.2 Localization experiment

Subjects were asked to localize sound sources when in the same room locations used for KEMAR measurements (Chapter 4 and Kopčo et al., 2001). Six normal-hearing subjects pointed to the perceived source location (five 150ms-long pink-noise bursts) presented from random locations between 0° – 90° azimuth and 0.15 – 1 m distance in the horizontal plane containing the ears. Each subject performed 300 trials in each room location. The (signed) mean error (re. actual source position) and standard deviation in response was computed from these results.

5.3 Results

5.3.1 Effect of reverberation on spectral cues

Figure 5-1 compares HRTF magnitude spectra at the four extreme source positions with KEMAR in the center of the room for anechoic and reverberant conditions. Reverberation adds frequency-to-frequency variability to magnitude spectra. This variability grows with source distance and is greatest at high frequencies. Variability increases with source azimuth for the ear contralateral to the source position and decreases with azimuth for the ipsilateral ear. Reverberation also fills in high-frequency notches, particularly at the far ear.

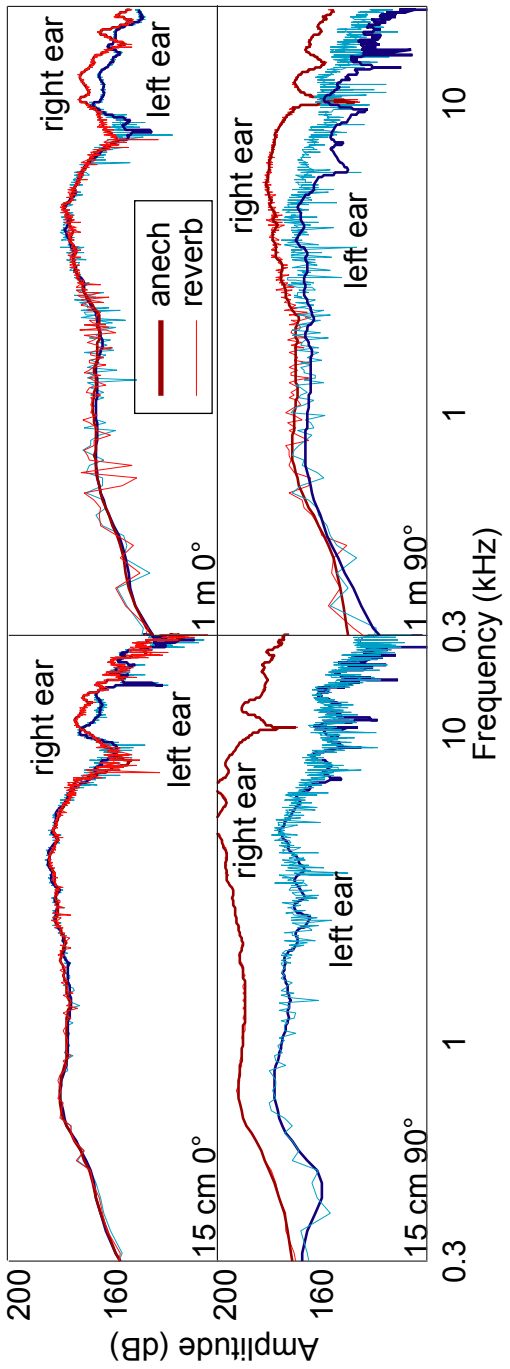


Figure 5-1 Anechoic and reverberant magnitude spectra at four source positions with KEMAR in center of room.

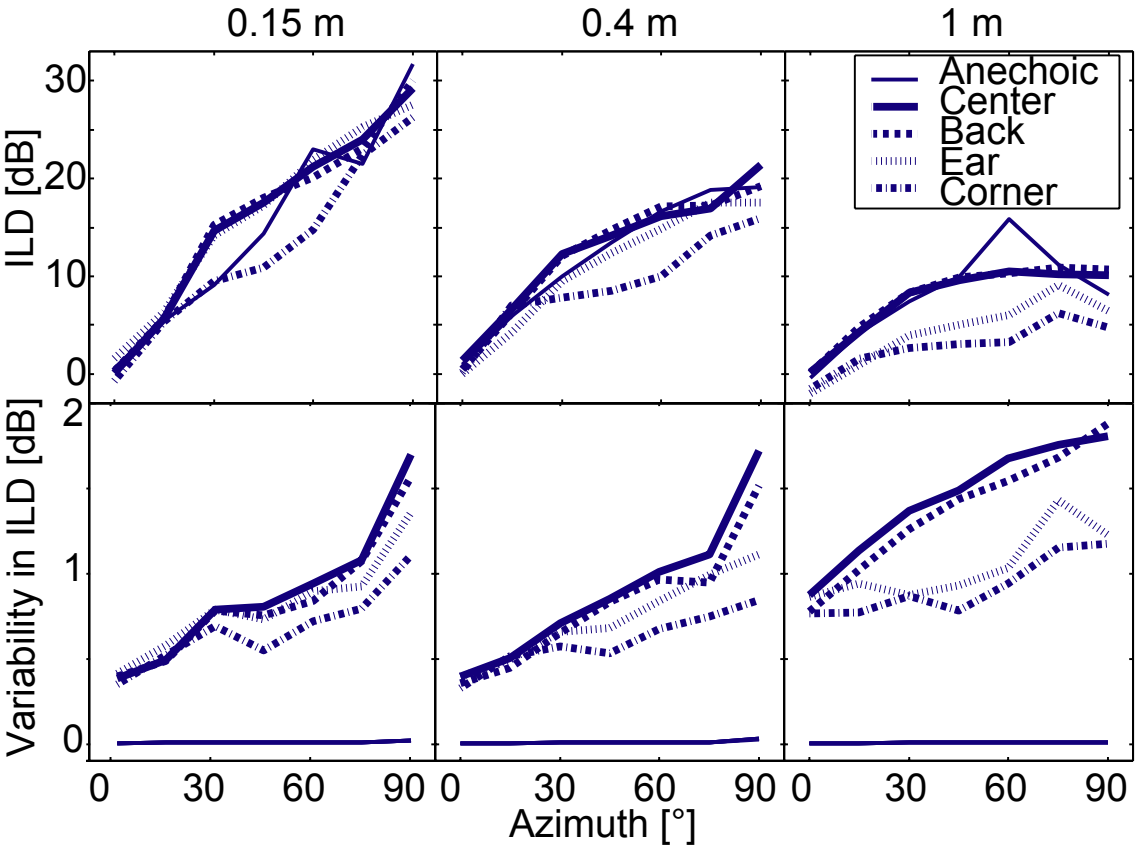


Figure 5-2 ILDs and cross-frequency variability in ILDs at 4 room locations as a function of source azimuth.

5.3.2 Effect of reverberation on ILDs

Figure 5-2 shows the ILD for different room locations and source positions. ILD magnitudes tend to decrease with reverberation, particularly for distant sources and conditions in which there is asymmetry in early reflections (Ear and Corner conditions). The frequency-to-frequency variability in the ILD (which is essentially zero in the anechoic condition) tends to increase with distance and is greatest for the Center condition. For room locations with early reflections, ILD variations are smoother and more systematic with frequency.

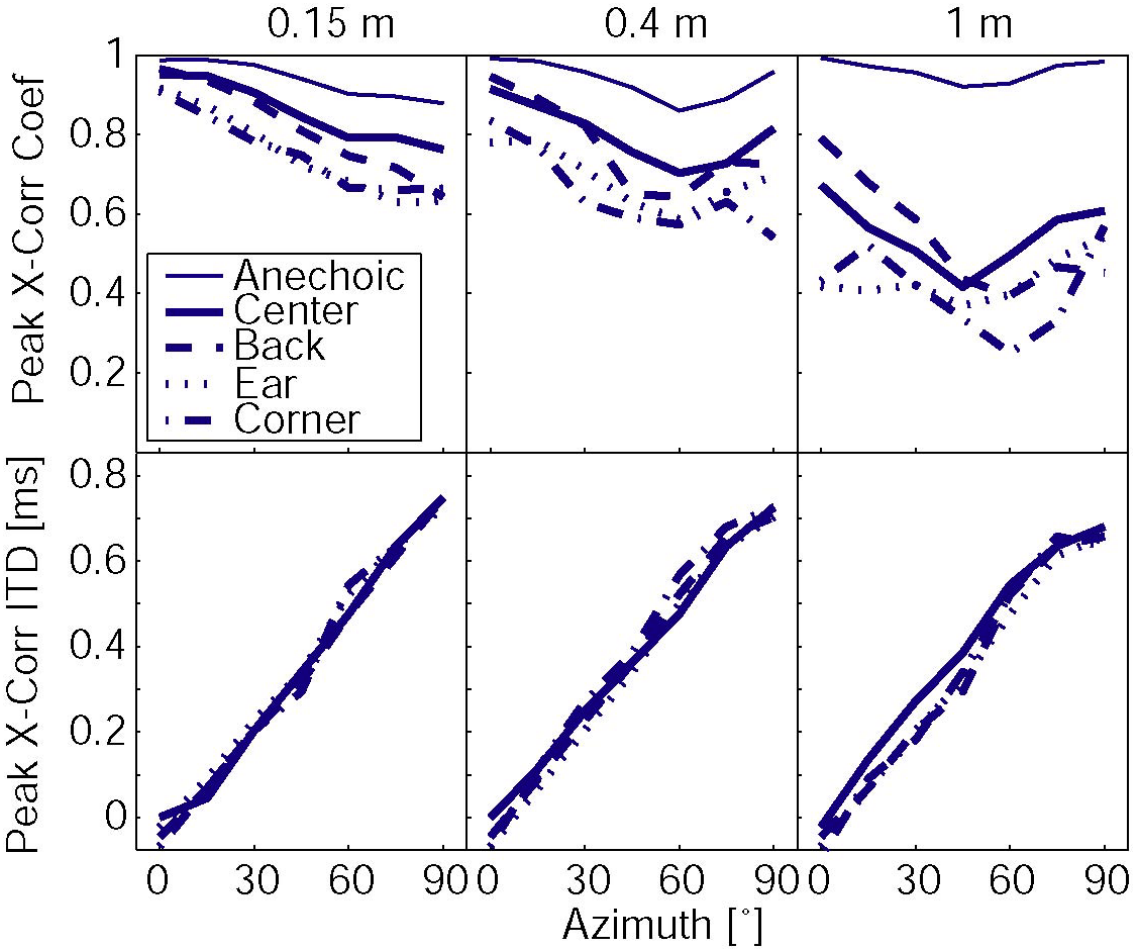


Figure 5-3 The peak value in the cross-correlation function within +/- 1 ms range and the corresponding ITD.

5.3.3 Effect of reverberation on ITDs

Within the biologically-plausible range (~0.8 ms), the ITD of the cross-correlation peak is roughly independent of source distance and room position (Figure 5-3). However, in reverberant conditions, the magnitude of this peak value decreases dramatically with distance and with the number of nearby walls. In addition, in the Corner and Ear conditions, a secondary peak (outside the biologically-plausible range of ITDs) can be of equal or larger magnitude than the primary peak in the cross-correlation.

Figure 5-4 illustrates that, as with the ILD, reverberation causes frequency-to-frequency variation in ITD. In the Center and Back conditions, this variation is essentially random around the "true" (anechoic) ITD. In the other conditions, the departures are more significant due to the early, asymmetric, strong reflections.

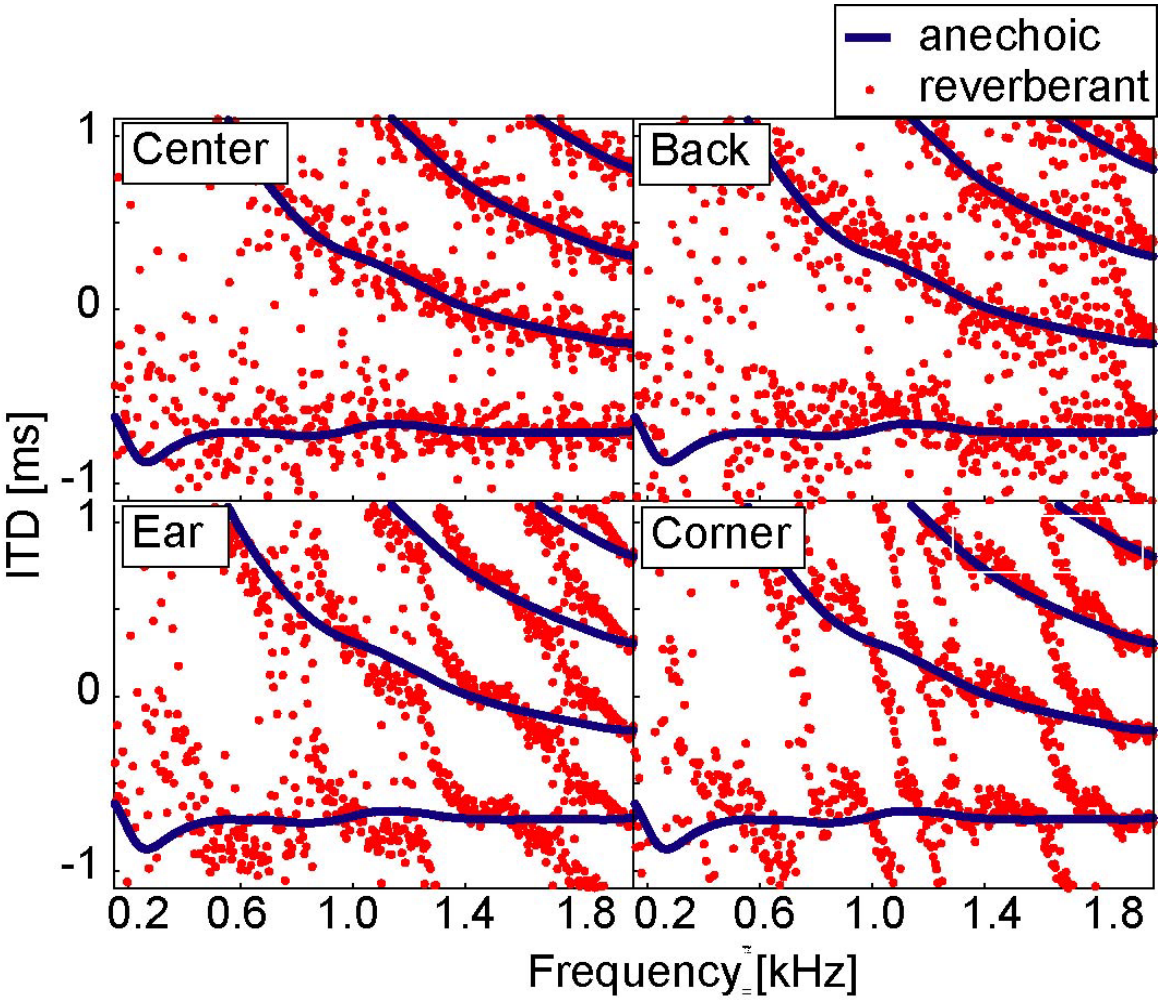


Figure 5-4 ITDs as a function of frequency in the anechoic and reverberant conditions for source at 90° 1 m.

5.3.4 Predictions vs. localization performance

Acoustic analysis shows that all localization cues in the signals reaching a listener are influenced by reverberation in a manner that depends on room position. To the extent that these cues determine spatial auditory perception, localization performance should also be influenced in a way that varies with listener location. Figure 5-5 summarizes behavioral results from a localization experiment performed in the room in which acoustic measurements were taken (Kopčo et al., 2001). Two small but statistically-significant trends were observed. 1) Azimuthal perception in the Back and Corner positions was biased towards the median plane (approximately 3.5°). 2) The variance in perceived azimuth was smallest for listeners in the Center condition, greatest in the Corner condition, and intermediate for the other two conditions (bottom row of Fig. 6).

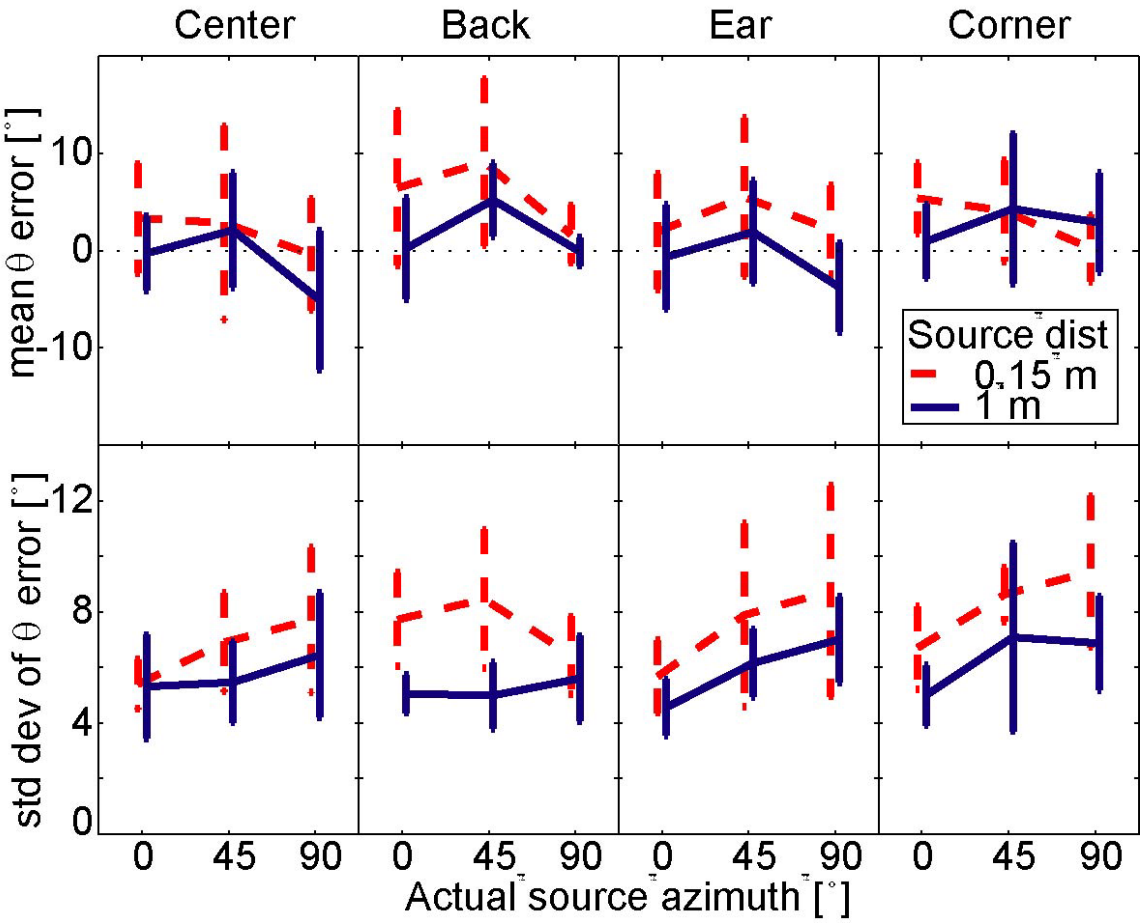


Figure 5-5 Across-subject mean and std. dev. of the response error, i.e., the difference between perceived and actual source azimuth.

The azimuthal bias is difficult to explain from results of the acoustic analysis. Acoustically, Ear and Corner conditions are most similar and most influenced by reverberation, but bias is only significant for Back and Corner locations.

On the other hand, the increase in the azimuthal response variance is consistent with both ILD variability and ITD decorrelation, which are greatest for the Ear and Corner conditions. This explanation cannot account for changes in bias with distance: the variability in acoustic parameters increases with distance while variance in perceived azimuth decreases with distance. The decrease in response bias with distance may be partially explained by the measurement method. If one assumes that response variability is constant in x-y-z coordinates, the same error translates to larger angular errors for nearby sources.

5.4 Summary and discussion

Acoustic analysis shows that the effect of reverberation on localization cues varies dramatically with listener position in a room. On the other hand, effects of room position on localization performance are modest, at best. Some of this apparent

discrepancy may be resolved by considering how acoustic cues change over time (as the current analysis evaluates only the expected value of the cues, ignoring variation in these cues over time). In fact, such dynamics are known to be perceptually important (cf. the "precedence effect"); for instance, the localization cues available at the onset of the stimulus will be much less distorted by reverberation than this first-order steady-state analysis suggests. Further, listeners may crudely estimate the effect of reverberation on the received stimuli and adjust the computation of source position accordingly. Future analysis will incorporate physiologically-based models of auditory processing (e.g., Colburn, 1977a) to predict how basic localization cues in reverberant signals may be extracted by the brain.

Chapter 6 PointMap: A real-time memory-based learning system with on-line and post-training pruning

Abstract

A new memory-based learning system, called PointMap, is introduced. PointMap, an extension of the Condensed Nearest Neighbor algorithm, evaluates the information value of its coding nodes during training, and uses this information to prune non-informative nodes both on-line and after training. These pruning methods allow PointMap to control both code size and sensitivity to detail in the training data. Pruning helps solve two problems of traditional memory-based learning systems: large memory requirements and sensitivity to noise. PointMap also overcomes the curse of dimensionality by considering multiple nearest neighbors during testing without increasing the complexity of the training process or the stored code. Information-value-based pruning can also be used in conjunction with other learning systems. The performance of PointMap is compared to performance of a group of 16 nearest-neighbor systems on several benchmark problems. Its performance is shown to be at least as good as performance of these algorithms, often approaching the Bayesian optimum.

Keywords: memory-based learning, nearest neighbor, on-line pruning, post-training pruning, incremental learning

6.1 Introduction

The nearest neighbor (NN) algorithm (Cover & Hart, 1967) is the basic algorithm of a group referred to as *memory-based learning systems*. During learning, NN forms a *code* by remembering all the training inputs and their associated outputs. During testing, NN finds, for a given unknown input, the most similar input(s) stored in the code, and assigns the unknown input to the same class. This strategy has been shown to achieve high classification accuracy while suffering from large memory and computation requirements, and sensitivity to noise (Dasarathy, 1991). Many methods have been proposed for minimizing these shortcomings. The basic approach is instance pruning (elimination of some instances from the code) (Lam et al., 2002). Instance pruning can be either incremental (starting with empty code and adding certain nodes) or decremental (starting with the complete code and removing useless nodes). (For a recent review see Wilson and Martinez, 2000). The pruning can prefer border points, central points, or points in between. Finally, the performance of different systems is typically evaluated in terms of the classification accuracy vs. the memory requirements and computational complexity (Lam et al., 2002).

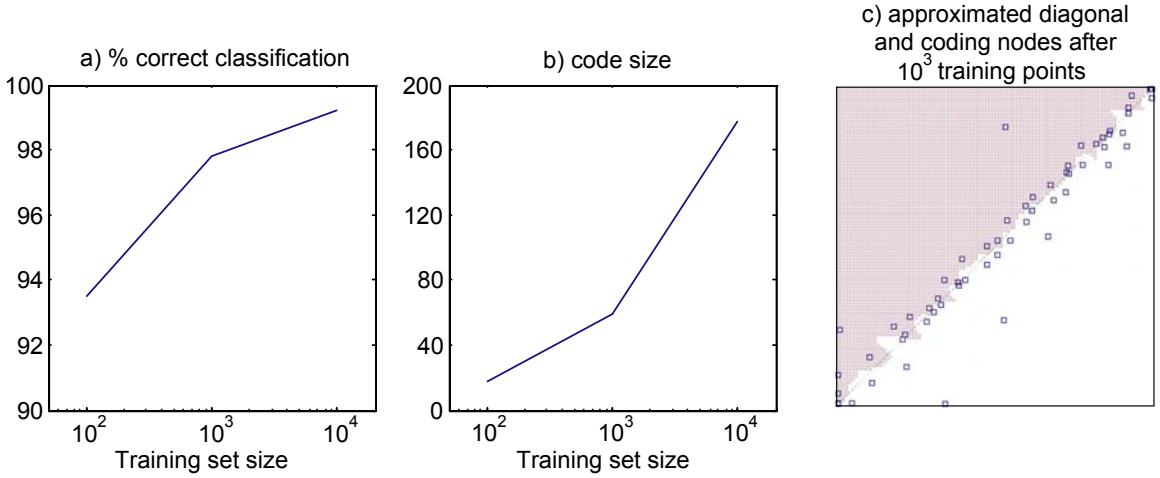


Figure 6-1 CNN algorithm simulations on the DIAGONAL data set. The training set consists of points uniformly distributed in the unit square, each labeled as lying above or below the diagonal. (a) Test set accuracy as a function of training set size. (b) Number of coding nodes as a function of training set size. (c) Test set response pattern (dark-above / light-below) after training with 10^3 points. Squares show the location of the 58 coding points.

The present study focuses on the methods of pruning in NN algorithms. The standard definition of the problem of a NN pruning method design is: given a training set T, find a subset S of T such that S is as small as possible while the classification accuracy based on S is at least as good as the accuracy based on T. This definition assumes that the whole set T is available at any point in time during training and that in principle it does not matter whether S is derived from T in incremental, decremental, or some kind of batch mode. The current study looks at the NN as a system with limited resources in a dynamic environment. That is, the system functions in an on-line mode with the inputs presented sequentially, one at a time, and the system has to learn to classify them without growing arbitrarily large. The criterion for the design of a pruning method for such NN system is:

"Let the maximum code size be fixed and assume that the system reached this maximum and it needs to add a new node. Find an optimum pruning strategy that will eliminate one of the current coding nodes such that, on average, the performance of the system with the pruned node replaced by the new one will improve."

The requirement that the performance must on average improve is necessary, because otherwise the system could stop learning once it reaches the maximum instead of adding a new node. Such pruning method is called *on-line* because the pruning decision must be made based only on the current state of the system and the current input. The problem of on-line pruning can be illustrated on the Condensed Nearest Neighbor (CNN, Hart, 1968) algorithm, the basic NN algorithm with incremental pruning. The algorithm starts with a coding set S containing one input from each class in T, and it adds new inputs from T to S only if they are misclassified by their nearest neighbor(s) in S. As shown in Figure 6-1, the accuracy of the CNN classification improves with presentation of more training points, however, the size of the code grows as well. The purpose of an on-line pruning method is to enable such a system to improve classification accuracy with presentation of more training points without the increase in the code size.

This chapter develops an incremental learning method that improves classification accuracy on-line while maintaining a fixed code size in memory-based systems. An on-line pruning method defines an indicator of the *information value* of coded items in a memory-based learning system called *PointMap*. PointMap estimates the information value of each stored input. Then, when new input needs to be added to the code, PointMap prunes the least informative node. Note that the addition of a new item may alter the information values of other coded items: in Figure 1c, for example, a point far from the decision boundary that is initially informative may lose its value as later points are added. In addition, information values may also be used to designate nodes for post-training pruning. Both on-line and post-training pruning help the learning system to find an informative subset of coding inputs among a redundant or noisy set.

Finding a simple local on-line pruning rule turns out to be non-trivial. Development of PointMap showed that the main feature of such rule is that it has to be conservative, preferring nodes that were in the system longer over the recently added nodes.

The resulting system has several advantages also from the point of view of the traditional, off-line learning. First, when learning is off-line the system can be thought of

as a whole class of learning systems, because the user can choose the maximum code size and the system will try to find the most accurate code with that size limit. So the choice of whether to achieve better code compression with lower accuracy or worse code compression with higher accuracy is on the user's side. As with other incremental pruning mechanisms, the on-line pruning enables the system to limit its code size already during training, i.e., it does not need to store the whole training set into memory at any point in time. Other advantages include: the ability to control whether the system should be sensitive to detail in the training data or whether it should try to generalize (i.e., whether it should retain inputs closer or farther from the boundary), the ability to identify and prune noisy data points, and the ability to completely re-learn the stored code if the system is in a non-stationary environment.

A post-training pruning mechanism can be added that further extends these advantages of the on-line pruning. For example, the on-line pruning allows the user to use cross-validation to determine the minimum code size necessary for accurate encoding of the training data. With post-training pruning, this size can be estimated without the need to repeat the learning process from scratch for every new maximum code size.

An important feature of the proposed system is its low complexity: the learning time is approximately the same as in the CNN, while the storage requirements can be arbitrarily low.

Two data sets from Wilson and Martinez (2000) were chosen to evaluate PointMap against sixteen nearest-neighbor systems with instance pruning. The first one, called WINE, is a small data set that was chosen with expectation that PointMap will not perform very well because it is more complex than basic NN systems and the data set is not large enough to allow it to estimate well its parameters. The second data set, called LED, was chosen because it was large, thus allowing PointMap to get a good estimate of each node's information value, and because it included multiple uninformative dimensions, thus testing PointMap's susceptibility to the curse of dimensionality. In both simulations PointMap achieved performance near the Bayesian optimum, proving that increase in the number of nearest neighbors considered during testing can overcome this curse on certain data sets.

Section 6.2 describes the PointMap algorithm in four parts: the CNN algorithm component (Section 6.2.1), the information value computation (Section 6.2.2), the on-line and post-training pruning computations (Sections 6.2.3-4), and the k-nearest-neighbor search used during testing (Section 6.2.5), followed by algorithm summaries for training and testing (Sections 6.2.6-7). Section 6.3 analyzes the behavior of the PointMap algorithm on benchmark problems, and compares its performance with that of other memory-based systems.

6.2 PointMap algorithm

PointMap is a memory-based learning system in the family of algorithms that include k-nearest-neighbor (kNN, Cover & Hart, 1967), Condensed-nearest-neighbor (CNN, Hart, 1968), IB2 (Aha, Kibler, & Albert, 1991), and Grow and Learn (Alpaydin, 1997). The memory of each of these algorithms is represented as a set of *coding nodes* which store input / output vector pairs from the training set. A PointMap node also records the number of times it has been chosen as the nearest neighbor of a new input, the

number of times it has made an incorrect prediction, and the number of times it has been deemed *critical* to the decision. A node that makes a correct prediction is defined as critical with respect to a given training input if the same system without that node would have made an incorrect prediction. These statistics are combined to estimate the *information value* of each node.

The PointMap training algorithm includes condensed nearest-neighbor selection of the coding node, update of the information value of that node and on-line and post-training pruning of nodes with low information value. During testing, PointMap performs a standard k-nearest-neighbor search, assigning each test input to the output class of the largest number of nodes among its k nearest neighbors.

6.2.1 Condensed nearest neighbor algorithm

The CNN algorithm performs an on-line search for a minimal predictive subset of the training data set. During training, the algorithm sequentially checks whether an input would be classified correctly by the candidate nearest neighbor in the current coding set. If yes, the algorithm proceeds to update the statistics of the candidate node, but otherwise does not alter the stored memory. If the prediction is incorrect, the current input is added to the coded memory. Thus, only the input vectors that cause predictive error are stored. Compared to the full training set stored by the k-NN algorithm, the reduction in the size of the coding set produced by CNN can be dramatic.

6.2.2 Information value of coding nodes

Two performance measures, *criticality* and *predictive accuracy*, define the information value of PointMap coding nodes. Updating of these two values requires only local computations at the chosen nearest neighbor node. The criticality fraction is defined as the ratio of the number of times the existence of a node has been critical to correct predictions relative to the number of times it has won and made correct predictions. The predictive accuracy fraction is defined as the ratio of the number of times a node has made correct predictions relative to the number of times it has won the nearest-neighbor competition. The information value of the node is then defined as the convex combination:

$$\textit{information value} \;=\; \gamma * \textit{criticality} \;+\; (1-\gamma) * \;\textit{predictive accuracy}$$

where the *criticality parameter* $\gamma \in [0,1]$ indicates the contribution that the criticality fraction makes to the information value.

Geometrically, the predictive accuracy fraction is highest for nodes that are far from the estimated decision boundaries, because nearby inputs tend to be from the same output class, by definition. In contrast, the criticality fraction is highest for nodes near a decision boundary, where elimination of a chosen node is most likely to produce a different prediction. Thus values of the criticality parameter γ determine the average distance between coding nodes and decision boundaries. Setting $\gamma = 0$ allows predictive accuracy alone to determine the coding set, which therefore tends to lie away from the decision boundary. Setting $\gamma = 1$ produces coding sets which are clustered near decision

boundaries. With noise-free training data, a wide range of γ values can produce reasonable results. However, coding nodes that represent noisy training inputs often have a high criticality fraction, so setting $\gamma = 1$ would encode much noise. On the other hand, setting $\gamma = 0$ and choosing nodes based on predictive accuracy alone could produce a code representing one small portion of input space where exemplars of one output class happen to be dense. For a given problem, γ can be chosen by validation. The simulation examples below indicate that setting the criticality parameter $\gamma = 0.25$ is a reasonable *a priori* choice, balancing accuracy (75% weight) with criticality (25% weight) in pruning decisions.

6.2.3 On-line pruning

An on-line pruning mechanism enables a system to limit its code size during training. Other advantages include the ability to balance sensitivity to detail against generalization, elimination of noisy data points, and the possibility of relearning the entire stored code, when necessary, in a non-stationary environment. In PointMap, maximum code size is fixed at a hard limit, C_{max}. During training, a new input exemplar is added when its nearest neighbor in the code fails to predict the correct output class. Once the size of the stored code reaches C_{max}, the least informative node is pruned before the new exemplar is added.

Note that the information value of a coding node lies between 0 and 1. When a new node is added, its information value is set equal to 0. This initial value represents an important design choice. It implies that a recently added node is almost always the first eliminated to make room for a new node as soon as the next incorrect prediction occurs. This elimination occurs even if the recent node was not involved in making the error. In order to remain in the code for long, a new node must quickly prove useful to other inputs. This property helps stabilize an existing code.

6.2.4 Post-training pruning

At any point during on-line learning, a subset of stored nodes with the smallest information values could be pruned from the coding set. Simulations in Section 3.2 demonstrate that post-training pruning on the basis of final information values can substantially improve test set performance with noisy data. In this case, many nodes in the noisiest regions may be stored during on-line learning. Since these nodes usually have relatively low predictive accuracy, then tend to be eliminated first in post-training pruning.

6.2.5 k-nearest neighbor testing

The curse of dimensionality many memory-based learning systems (Cybenko et al., 1994). That is, when inputs are high-dimensional, these systems need to store many points for sufficiently dense coverage of the space, even though some of the dimensions may be irrelevant to outcome predictions. PointMap alleviates this problem by using multiple nearest neighbors to predict outcomes during testing. This strategy improves performance because each of the neighbors serves as an independent noisy version of the

current input. Voting among the neighbors is equivalent to comparing the input to the average of the neighbors from each category, with averaging effectively eliminating noise in irrelevant dimensions.

Table 6-1 PointMap variables

Description	Parameter
Training set index set	$p = 1 \ldots P$
Current input vector	$\mathbf{I} = \mathbf{I}_p \equiv \left(I_{p1} \ldots I_{pi} \ldots I_{pM} \right)$
Correct output class for the current input	$O = O_p$
Code index set (stored coding nodes)	$j = 1 \ldots C$
Code vector j	$\mathbf{w}_j = \left(w_{j1} \ldots w_{jM} \right)$
Output class associated with coding node j	Ω_j
Index of the coding node closest to the current input	J
Index of the next closest coding node	J^{next}
Index of coding node for pruning	J^{prune}
Number of inputs for which coding node j won	α_j
Number of inputs for which coding node j won and made a correct prediction	β_j
Number of inputs for which coding node j won, made a correct prediction, and was critical	χ_j
Information value of coding node j	δ_j
Index set of the k nearest neighbors during testing	$\lambda \in \Lambda$

Table 6-2 PointMap parameters

Description	Parameter
Criticality parameter	$\gamma \in [0,1]$
Maximum code size	$C_{max} \in [1,\infty]$
Number of nearest neighbors used for testing	$k \in [1,C]$
Fraction of nodes to be retained after post-training pruning	$\theta \in [0,1]$

6.2.6 PointMap training algorithm

PointMap is trained on P input-output pairs $(\mathbf{I}_1,O_1)$, $(\mathbf{I}_2,O_2)$, … , $(\mathbf{I}_P,O_P)$. The input vector $\mathbf{I}_p$ has M components $(I_{p1}\ldots I_{pi}\ldots I_{pM})$, and the integer value O_p represents the output class to which $\mathbf{I}_p$ belongs. The stored code consists of the input-output pairs $(\mathbf{w}_1,\Omega_1)$, … , $(\mathbf{w}_C,\Omega_C)$.

Table 6-1 lists the variables used in the PointMap algorithm. In addition, four parameters are determined by the user (Table 6-2). Three of these parameters influence PointMap training: criticality γ biases the system toward choosing coding nodes that are farther from $(\gamma = 0)$ or closer to $(\gamma = 1)$ estimated decision boundaries; code size C_{max} sets an upper bound on the size of the coding set; and a post-training pruning fraction θ determines how many coding nodes with low information values are discarded before testing. For testing, parameter k specifies the number of neighbors that vote on the output prediction.

Individual steps of PointMap training are defined as follows. A Matlab implementation of the PointMap code with a sample demo may be found at: http://cns.bu.edu/~pointmap.

Step 1: Code the first input

Set $p = C = 1$

$\mathbf{w}_1 = \mathbf{I}_1$

$\Omega_1 = O_1$

$\alpha_1 = \beta_1 = \chi_1 = \delta_1 = 0$

Step 2: Present a new input Vector **I** denotes the current training input, and O is its output class.

Increase p by 1

Set $\mathbf{I} = \mathbf{I}_p$

$O = O_p$

Step 3: Choose a candidate coding node J The algorithm searches the stored code for the nearest neighbor of **I** using the L_1 (city-block) metric.

$$J = \arg \min_{1 \le j \le C} \left| \mathbf{I} - \mathbf{w}_j \right|,$$

where $\left| \mathbf{I} - \mathbf{w}_j \right| \equiv \sum_{i=1}^{M} \left| I_i - w_{ji} \right|$, with ties broken in favor of the smallest index.

Step 4: Update the information value of the candidate node J

Update the number of times α_J that node J has won the nearest-neighbor search:

Increase α_J by 1

If $O_J = O$, update the number of times β_J that node J has produced a correct prediction and the number of times χ_J that node J has been critical to the correct prediction:

Increase β_J by 1

Let $J^{next} = \arg \min_{\substack{1 \le j \le C \\ j \ne J}} \left| \mathbf{I} - \mathbf{w}_j \right|$

If $O_{J^{next}} \ne O$ (or if $C = 1$), increase χ_J by 1

Recompute the information value δ_J of the candidate node J:

$$\delta_J = \gamma \frac{\chi_J}{\beta_J + 1} + (1 - \gamma) \frac{\beta_J + 0.5}{\alpha_J + 1}$$

Step 5: If node J **has made the correct prediction, go to the next training item**

If $O_J = O$, go to Step 7

Step 6: If J has made an incorrect prediction, add a new coding node

If $C = C_{max}$, then eliminate the stored node J^{prune} with the smallest information value:

$J^{prune} = \arg \min_{1 \le j \le C_{max}} \delta_j$ (In case of a tie, choose the smallest index.)

For $j = \left(J^{prune} + 1\right) .. C_{max}$

$\mathbf{w}_{j-1} = \mathbf{w}_j$

$\Omega_{j-1} = \Omega_j$

$\alpha_{j-1} = \alpha_j$

$\beta_{j-1} = \beta_j$

$\chi_{j-1} = \chi_j$

$\delta_{j-1} = \delta_j$

Set $C = C_{max} - 1$

Initialize a new node that encodes the current input.

Increase C by 1

$\mathbf{w}_C = \mathbf{I}$

$\Omega_C = O$

$\alpha_C = \beta_C = \chi_C = \delta_C = 0$

Step 7: End condition The algorithm performs a single pass through the training set. If $p < P$, go to Step 2

Step 8: Post-training pruning Reduce the stored code to a fraction θ of its final on-line size.

Let $C_\theta = \theta * C$

While $C > C_\theta$, sequentially eliminate the stored nodes J^{prune} with the smallest information values:

$$J^{prune} = \arg \min_{1 \le j \le C} \delta_j$$

For $j = \left(J^{prune} + 1\right) .. C$

$$\mathbf{w}_{j-1} = \mathbf{w}_j$$

$$\Omega_{j-1} = \Omega_j$$

$$\delta_{j-1} = \delta_j$$

Reduce C by 1

6.2.7 PointMap testing algorithm

The predicted output class of a test input is determined by a majority vote of its k nearest neighbors in the stored code. The algorithm performs no further learning or pruning, and does not use information values during testing.

Step 1: Presentation of new input

Let **I** be the test input.

Step 2: Identify the k nearest neighbors in the stored code

Let $\Lambda \subseteq \{1 \ldots C\}$ be the set of k coding indices such that $\left| \mathbf{I} - \mathbf{w}_\lambda \right| \le \left| \mathbf{I} - \mathbf{w}_j \right|$ for all $\lambda \in \Lambda$ and $j \notin \Lambda$, with ties broken in favor of the smallest index.

Step 3: Output class prediction The predicted output class is determined by a vote among the k nearest neighbors of the test input. That is, O is taken to be the class with the largest representation in the set $\{\Omega_\lambda : \lambda \in \Lambda\}$. Tie votes may be broken in favor of the smallest output class number, or the output class of the nearest neighbor, or by weighting votes according to distances to the test input.

6.3 PointMap simulations

The PointMap algorithm was tested on four simulation examples, with performance compared to that of other memory-based learning systems. The first simulations (Section 3.1) examine PointMap behavior on the noise-free SPRING data set, which features a multi-scale decision boundary. The second simulations (Section 3.2) examine performance on the same task with a high degree of noise added to the training set. Sections 3.3 and 3.4 examine system performance on benchmark examples (WINE and LED) from the UCI repository of machine learning databases (Merz, 1996). PointMap results on the WINE and LED examples are compared to those of sixteen other k-NN-based learning systems, as analyzed by (Wilson and Martinez, 2000).

6.3.1 SPRING simulations

The SPRING benchmark was constructed to test whether PointMap had achieved its design goals. This section addresses the following questions on a noise-free version of the SPRING example:

Once the stored code has reached its size limit, does further training improve test performance?

Given a sufficiently large code size limit, can further training approach optimal performance?

Can post-training pruning be used to estimate how many coding nodes are necessary to accurately represent the data set? That is, if C_{max} is chosen to be larger than a minimum necessary for accurate performance, will post-training pruning be able to eliminate redundant nodes without loss of accuracy?

What is a good *a priori* parameter value to balance criticality γ vs. predictive accuracy $(1-\gamma)$?

6.3.1.1 SPRING data and simulation parameters

SPRING input points are uniformly distributed within the unit square. Points in the two output classes are located to the left or right of a zig-zag, multi-scale decision boundary (Figure 6-2). The present noise-free simulations have every point in the training set labeled correctly. The following simulations examine system performance with up to 4×10^6 training exemplars, three code size limits (C_{max} = 20, 50, 200), and four criticality parameters ($\gamma = 0$, 0.25, 0.5, 1). After each presentation of 8×10^4 training points, system performance was tested with three levels of post-training pruning (θ = 0.5 $(\cdots)$, 0.75 (---), 1.0 (—)) on a test set grid of 101×101 points. The nearest-neighbor parameter k was set equal to 1 during testing.

6.3.1.2 SPRING results

The solid lines in Figure 6-3 show how test set performance changes with additional training, after each network has reached its maximum size C_{max}, without post-training pruning $(\theta=1)$. For each criticality value γ (columns), the correct prediction rate increases with C_{max}, reaching close to 100% when $C_{max}=200$.

The rows of Figure 6-3, show that, for all network sizes, performance of the unpruned system is best when criticality contributes 25% of the information value. Moreover, for this value of γ, test set performance consistently improves with incremental learning for all code sizes. In addition, the dashed line in Figure 6-3j shows that, with $C_{max} = 200$, this system maintains test performance levels even after pruning the 50 coding nodes with lowest information values $(\theta = 0.75)$.

Note, however, that the more drastically pruned systems $(\theta = 0.5)$ maintain better performance with $\gamma = 0$. In this case, the information value is based on predictive accuracy alone, so the system is selecting for good generalization as opposed to finer details of the decision boundary.

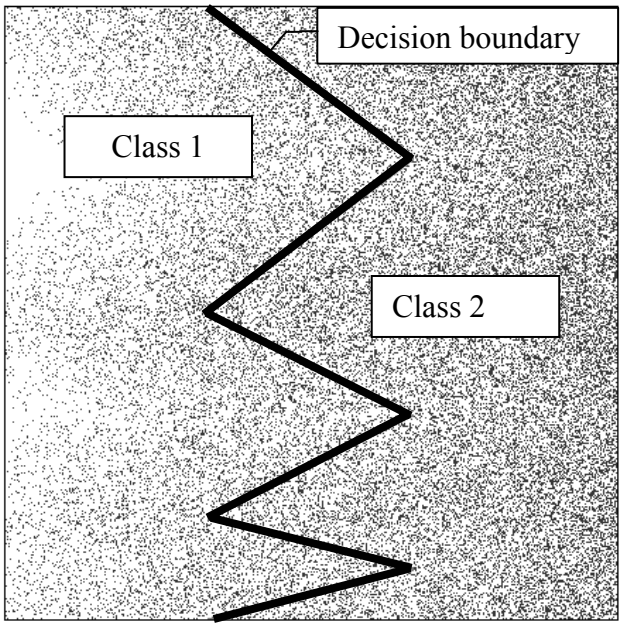


Figure 6-2 For the SPRING problem, a multi-scale zig-zag marks the ideal boundary between the two classes of points in the unit square. In the noisy version, the probability of training set point having the wrong label is proportional to its distance to the boundary. Points in the figure represent 10^5 training exemplars from Class 2 in a noisy SPRING example.

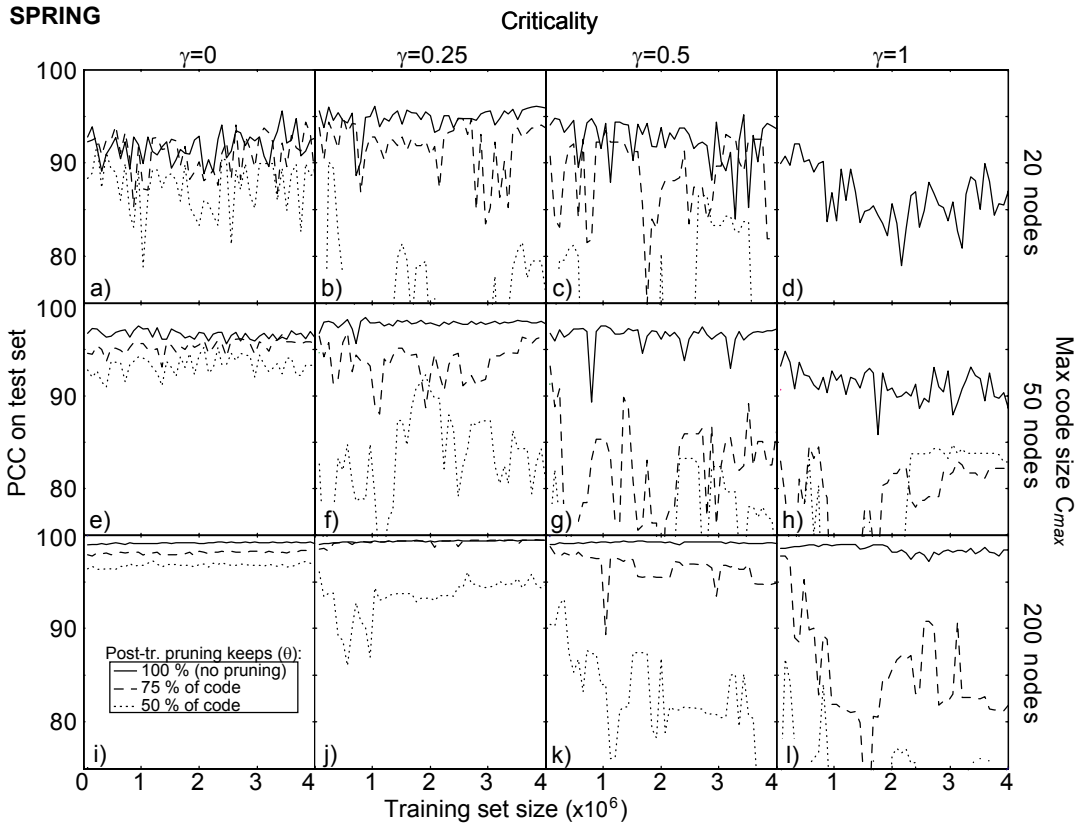


Figure 6-3 Simulations on the noise-free SPRING example illustrate the role of criticality $\gamma = 0.0, 0.25, 0.5, 1.0$ for computing the information value of each point, of the maximum code size $C_{max} = 20, 50, 200$, and of the post-training pruning fraction $\theta = 1.0, 0.75, 0.5$. When $\gamma = 0$, criticality does not contribute to the information value, which is based on predictive accuracy alone; when $\gamma = 1$, the information value is based only on criticality. In each column, performance is seen to improve as the maximum code size increases. Within each panel, the solid line shows how test set performance varies with the number of training points; the dashed line shows performance by the 75% of nodes with highest information values; and the dotted line shows performance by the top 50% of trained nodes. Setting $\gamma = 0.25$ achieves the best results.

Finally, when the information value is based on criticality alone $(\gamma = 1)$, performance levels are consistently worse than in systems with smaller γ values. For all code sizes, on-line pruning even causes performance to deteriorate with additional training. This last observation indicates that nodes stored early in training that provide some degree of generalization are gradually discarded in favor of nodes near decision boundaries. Choosing values of $\gamma < 1$ help avoid this sort of over-fitting.

Figure 6-4 illustrates PointMap dynamics for the simulations summarized in Figure 6-3. Here, each plot shows the decision regions from the beginning (after 8×10^4 inputs) and end (after 4×10^6 inputs) of training, as well as the locations of points in the stored code and the test-set percent correct. These graphs show that, although there is no explicit feedback among nodes, on-line pruning helps to create an evenly distributed coding set. The rows of Figure 6-4 also illustrate that the average distance from coding nodes to the decision boundary decreases as the criticality parameter γ increases. In the top row, where the code size is limited to $C_{max} = 20$, the system performs significantly better with small values of γ, which keeps coding nodes away from the decision boundary, although some contribution of criticality (Figure 6-4b) is better than none (Figure 6-4a).

Note that the average distance between coding nodes and decision boundary also decreases as the maximum code size C_{max} increases. When sufficiently many coding nodes are allowed, the system is able to distribute them near the decision boundary, to fine-tune accurate prediction across multiple scales. A comparison of Figure 6-4a with Figure 6-4i shows that this clustering near the boundary can occur even with $\gamma = 0$, which otherwise tends to place nodes as far as possible from the boundary, for generalization. Clustering near the boundary can nevertheless occur in this case because coding nodes are added only when the system makes a predictive error, which tends to occur near the boundary.

Figure 6-4e shows that, as code size increases, larger-scale sections of the decision boundary are accurately delineated, while smaller scale portions are approximated. Even with the same number of coding nodes, setting $\gamma = 0.25$ (Figure 6-4f) pulls the coding nodes toward the boundary, improving the approximation at the smaller spatial scales.

Finally, Figure 6-4d indicates why setting $\gamma = 1$ produces a poor approximation to the decision boundary. Relying on criticality alone favors the selection of pairs of coding nodes which are near one another across the boundary. Each has a high criticality factor when it makes a correct prediction for a training input, because its removal would produce the incorrect prediction. These tightly clustered pairs produce over fitting because small misalignment of the nodes can produce a large error in the approximated decision boundary. With $\gamma = 0.25$, many pairs of coding nodes still appear on opposites sides of the boundary, but at some distance, producing a more stable approximation.

6.3.2 Noisy SPRING

To evaluate PointMap's ability to cope with noise, simulations were performed on a noisy version of the SPRING example. Memory-based learning systems that look only for the nearest neighbor (1-NN or 1-CNN) would either perform poorly on this type of example, or they would generate a large code. The performance of these classifiers can be improved by increasing the number of the nearest neighbors (k) making test set predictions. Although larger values of k allow a system to estimate the class probabilities in the region around a test input, this computation increases the complexity of the algorithm and does not reduce the code size. Simulations below show that PointMap can

find a good solution (95% correct) to the noisy SPRING problem with a relatively small coding set and using $k = 1$ nearest neighbor during both training and testing.

6.3.2.1 Noisy SPRING data and simulation parameters

For the noisy SPRING data set, output class labels were randomly swapped on a subset of the training set. The probability of a class swap was 50% for points at the decision boundary, with the swapping probability decreasing with distance between input points and the boundary. Figure 6-2 shows 10^5 training points assigned to Class 2.

The following simulations examine PointMap performance with one code size limit $(C_{max} = 200)$, two criticality parameters ($\gamma = 0$, 0.5), and three levels of post-training pruning (θ = 0.1 $(\cdots)$, 0.5 (---), 1.0 (—)). As for the SPRING simulations, after each presentation of 8×10^4 training points, system performance was tested on a grid of 101×101 points, with k = 1 for nearest-neighbor search.

6.3.2.2 Noisy SPRING results

Figure 6-5 illustrates PointMap performance on the noisy SPRING example, with criticality parameter $\gamma = 0$ for the left column and $\gamma = 0.5$ for the right column. Figure 6-5a shows that, when the information value is based on predictive accuracy along $(\gamma = 0)$, test set performance reaches about 95%, and remains in that range even when post-training pruning reduces the stored code from 200 to 20 nodes $(\theta = 0.1)$.

Figure 6-4 Initial and final SPRING coding node distributions, for simulations of Figure 6-3, without post-training pruning $(\theta = 1)$. Each panel shows the predicted decision region and stored coding points after an initial training phase (8×10^4 inputs) and at the end of the simulation (4×10^6 inputs). These simulations show that, once a network has achieved its maximum size, additional training does not automatically improve performance. In fact, at each network size with $\gamma = 1$, on-line pruning pulls stored points closer to the decision boundary, but this additional training leads to a deterioration of test-set accuracy. In contrast, with $\gamma = 0.25$, on-line pruning improves accuracy at each network size.

Figure 6-5 Noisy SPRING simulations with $C_{max} = 200$. The information value in the left column is based on predictive accuracy alone $(\gamma = 0)$ and in the right column is equally weighted between predictive accuracy and criticality $(\gamma = 0.5)$. (a) System performance as a function of training set size, averaged across five simulations. (b) Final distribution of coding nodes and decision regions with no post-training pruning. (c) Same as (b), retaining 50% of the trained nodes. (d) Same as (b), retaining 10% of the trained nodes.

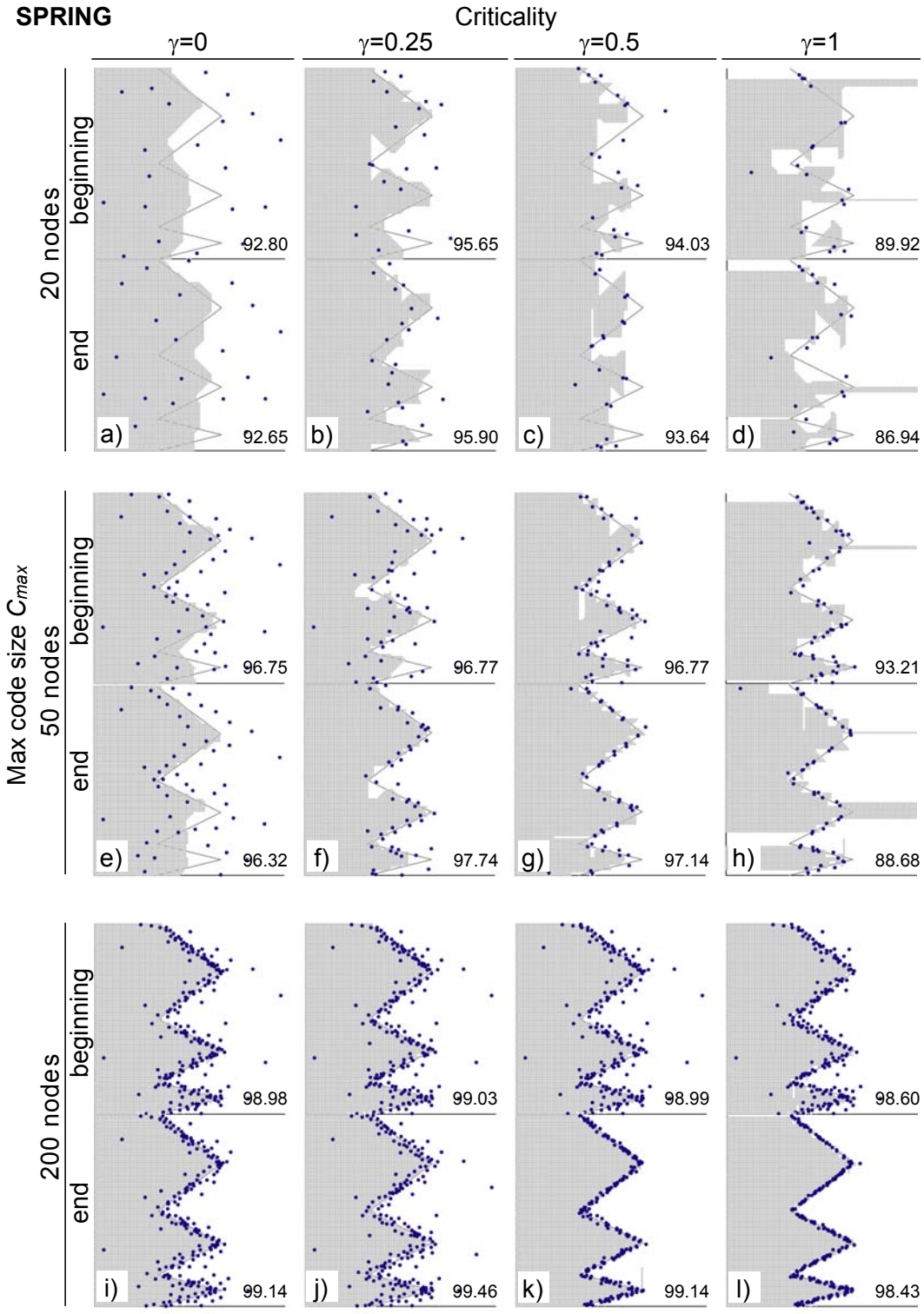


Figure 6-4 (see caption on previous page)

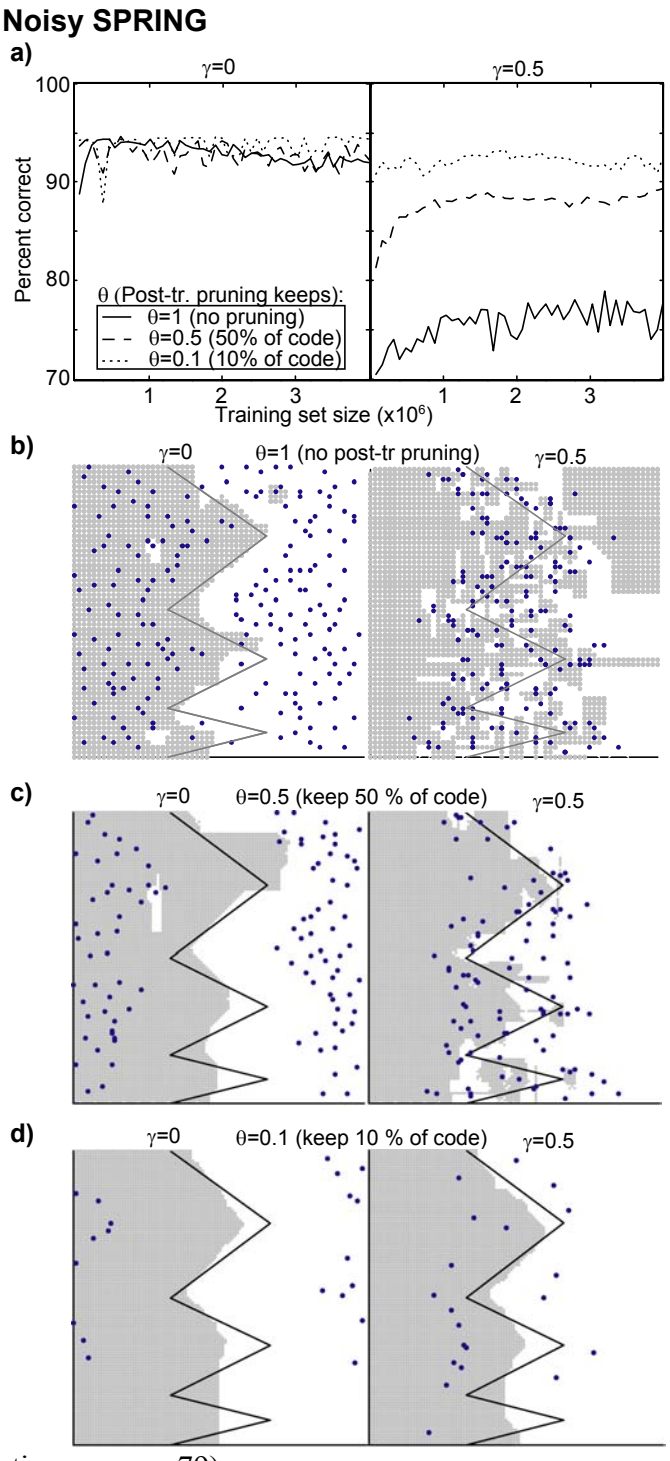


Figure 6-5 (see caption on page 79)

At first glance, this performance appears superior to that of the system with $\gamma = 0.5$, especially without post-training pruning $(\theta = 1)$. Closer inspection reveals additional subtleties in this comparison. With $\gamma = 0$, after an initial stable phase, performance steadily deteriorates with additional training. In contrast, with $\gamma = 0.5$, performance steadily improves, especially with post-training pruning.

Figure 6-5b-d, which shows the final stored code and test set decision region for each criticality setting and each degree of post-training pruning, provides insight into these dynamics. Coding nodes selected on the basis of their predictive accuracy alone (left column) tend to be located at some distance from the actual decision boundary. The bias toward placing the code as far as possible from the decision boundary is balanced by the fact that a new node is added only in response to a predictive error. However, Figure 6-5c-d shows that it is the nodes located farthest from the decision boundary, which have information value is close to 1, that remain after pruning. Although test set performance remains high for this particular example, the potential for over-generalization with $\gamma = 0$ is visible where the pruned decision boundary fails to approximate the higher frequency portion of the spring. Slow migration of the code away from the decision boundary, and resulting over-generalization, also explains the deteriorating performance in the course of on-line learning when criticality is not a factor in the information value

The right column of Figure 6-5b shows that giving equal weight to criticality and predictive accuracy $(\gamma = 0.5)$ clusters coding nodes in regions with a high concentration of noise, thus producing a poorly defined test set decision boundary. However, because information values are higher away from the actually decision boundary, test set performance of the pruned system steadily improves during training, reaching a level equal to that of the best performance with $\gamma = 0$. Moreover, pruning improves the geometry of the decision boundary approximation.

These simulations hereby indicate that a problem with a high degree of noise is best approached by a coding strategy that balances criticality and predictive accuracy during on-line training, followed by post-training pruning. However, validation set selection of the free parameters γ and θ may be time-consuming. The *a priori* parameter selection $\gamma = 0$ and $\theta = 1$, which chooses nodes on the basis of predictive accuracy alone without post-training pruning, provides a quicker if less compact solution.

6.3.3 WINE simulations

The previous sections analyzed the properties of PointMap on synthetic data sets. The following two sections compare the performance of PointMap to performance of sixteen variations of the *3*-nearest-neighbors classifier analyzed by (Wilson and Martinez, 2000). The present section compares these systems on the WINE data set, which features a small training set. This example was chosen specifically to challenge PointMap, which was designed primarily to estimate coding node information values from large training sets.

6.3.3.1 WINE data and simulation parameters

The WINE data set was obtained from chemical analysis of wines grown in a single region in Italy, derived from three vine varieties. The set consists of 178 13-dimensional input patterns, each belonging to one of three output classes. Although the classes are linearly separable, memory-based learning systems usually do not find the optimum solution because of the small number of available training inputs. PointMap simulations employed the same 10-fold cross validation design used in the Wilson-Martinez analysis. That is, each of ten partitions of the data set served, in turn, as a test set, with training on the remaining set of ~160 items and results reporting the average across ten the trials.

Based on estimates obtained from a preliminary pilot study, the PointMap criticality parameter was fixed at $\gamma = 0.15$ across all simulations. Testing employed $k = 1$ nearest neighbor. Each training set consisted of 500 random permutations of the training set. Typically, the stored code stabilized early, but later epochs improved estimation of information values, for post-training pruning. Simulations examined maximum network sizes C_{max} ranging from 3 to 40, and post-training pruning fractions $\theta = 0.25, 0.5, 0.75, 0.9, 1.0$. An additional post-training pruning strategy that eliminated only the last-created node $(\theta = 0.999)$ was also tested.

6.3.3.2 WINE results

Figure 6-6plots system performance (percent correct) on the WINE example as a function of the average number of input vectors stored in memory. Plotted points summarize performance of the 16 variations of the *3*-NN classifier reported by (Wilson and Martinez, 2000). These systems produced only small variations in percent correct, averaging 93.5% and with all but one between 90.98% and 96.08%. However, the size of the stored memories ranged from ~3 nodes (81.47% correct) to all the nodes for *3*-NN (94.93% correct).

The solid curve in Figure 6-6 plots an exponential fit of PointMap performance. Although the maximum code sizes up to $C_{max} = 40$ were tested, actual code sizes never exceeded $C = 23$. Therefore simlations with larger C_{max} values used no on-line pruning, in which case variations were caused only by differing orders of input presentations. Although PointMap uses only $k = 1$ nearest neighbor during testing, performance is similar to that of the collective Wilson-Martinez systems for at each code size. With pruning of the last-created node (widest dashes), PointMap performance improves further at small code sizes, with the system discovering a near-optimal solution with only $C = 3$ coding nodes, and maintains performance with larger stored codes. The PointMap on-line pruning strategy is thus seen to succeed with small data sets as well as large ones.

Additional curves in Figure 6-6 show exponential fits of PointMap performance on WINE simulations following pruning that retains from 90% (dashes) to 25% (dots) of the trained code. These results indicate that post-training pruning causes rapid performance deterioration, even when many more nodes are retained than the minimum needed for optimal performance. This example therefore suggests an *a priori* strategy for small training sets that prunes on line after reaching a small upper bound on code size and that

discards only the last node added during training. Note that this strategy differs from the one suggested for large noisy data sets, which was to prune a fraction of a larger code after training.

6.3.4 LED simulations

The final simulations compare PointMap performance with that of the same 16 systems as in Section 3.3, again as reported by (Wilson and Martinez, 2000). Compared to the WINE example, this LED benchmark has more input dimensions, output classes, and training exemplars, and intrinsic noise establishes an upper bound on test set performance.

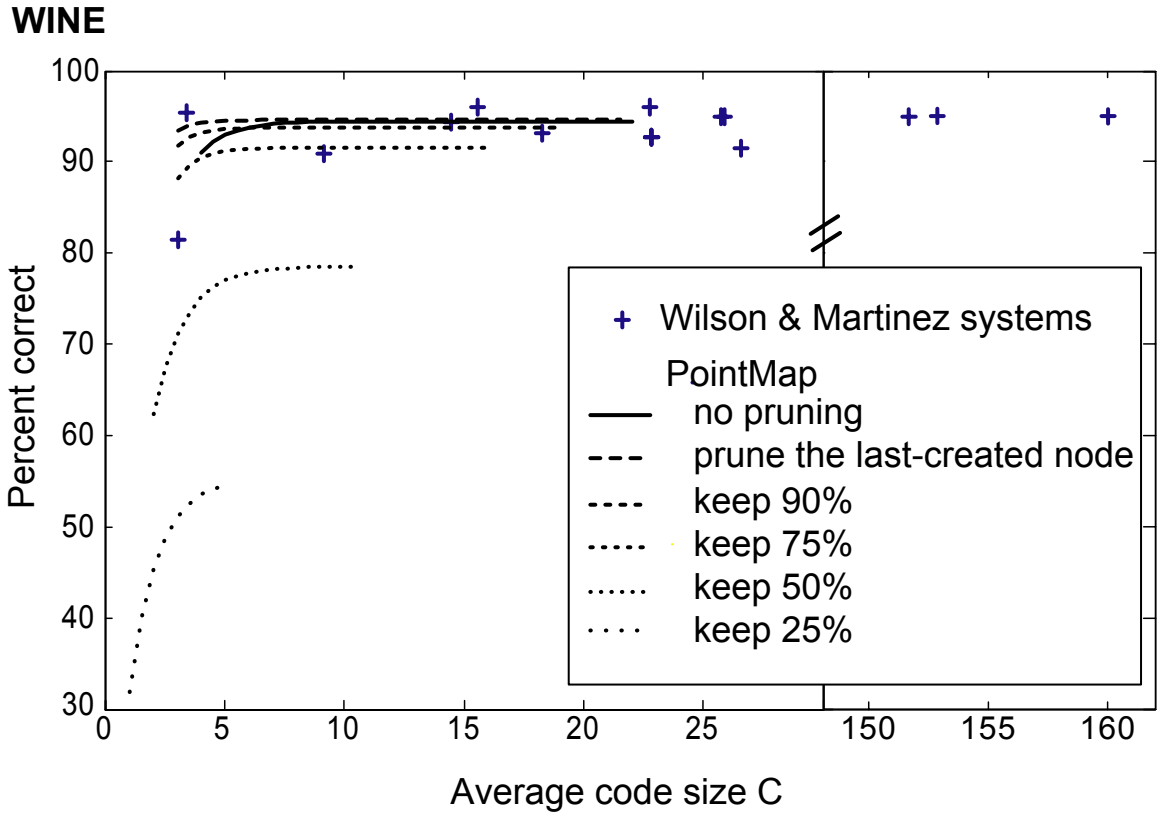


Figure 6-6 WINE simulations with $\gamma = 0.15$. For each system, classification accuracy, as a function of memory size, is averaged across 10-fold cross validation trials. Crosses denote the 16 *3*-NN classifiers reported by (Wilson and Martinez, 2000). PointMap results are plotted by exponential fit for the unpruned system (solid) curve, with progressively shorter dashs marking increasing levels of post-training pruning.

6.3.4.1 LED data and simulation parameters

The LED task is to identify numbers 0,1,…,9 in a digital display. The first seven components of the binary input vector denote the presence or absence of a segment in an ordinary display. The task is rendered difficult by the presence of 17 more random input dimensions. In addition, values in the first seven components are flipped with a 10% probability. Because some flips can change one digit into another (e.g., **6** into **8**), the optimal Bayes classification rate is 74%. Each output class is represented by 1,000 input vectors.

In PointMap simulations, each training set was presented in random orders for 100 epochs. As in the Wilson-Martinez paradigm, performance represents averages from 10-fold cross validation. In a preliminary simulation (Section 3.4.2), the criticality parameter γ was set to 0, the maximum code size $C_{max} = 40$, and the number of nearest neighbors $k = 1$, with no post-training pruning. The main simulation (Section 3.4.4) set $\gamma = 0.25$, with the maximum code size C_{max} = 30, 100, 300, 1000, and 3000 nodes, with no post-training pruning, and with parameter k chosen by 10-fold cross-validation on the training set.

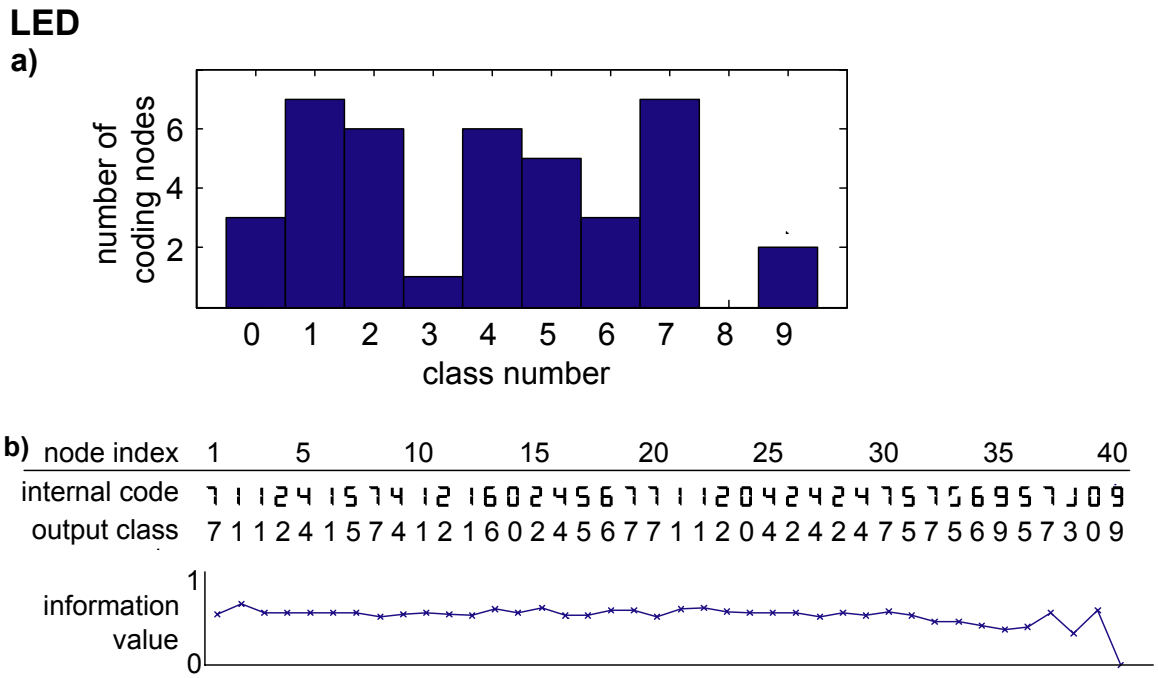


Figure 6-7 Sample PointMap simulation of the LED example with γ=0, $C_{max} = 40$, 100 epochs, and no post-training pruning $(\theta = 1)$. (a) Histogram of the number of coding nodes for each class at the end of training. (b) For each of the 40 stored nodes: its index (with recently created nodes having larger indices), its internal code (from input components #1-7), the class to which it is assigned, and the final estimate of its information value.

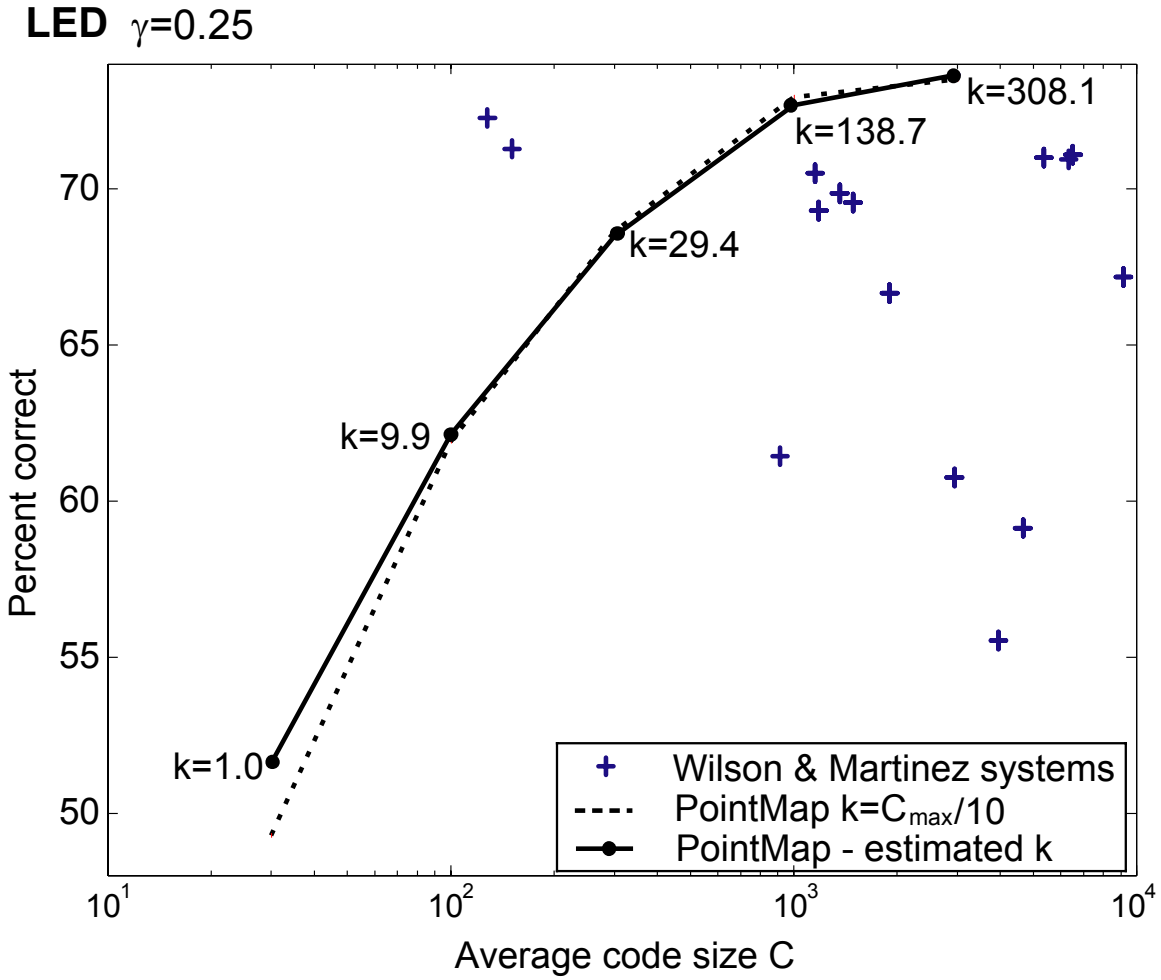


Figure 6-8 LED simulation with $\gamma = 0.25$ and no post-training pruning. Crosses mark results from the sixteen $3-NN$ systems reported by **(Wilson and Martinez, 2000)**. The solid line shows average PointMap performance for $C_{max} = 30$, 100, 300, 1,000, and 3,000 nodes, the number of test set nearest neighbors k determined by 10-fold cross-validation on the training set. The dotted line shows PointMap results with $k = C_{max}/10$.

6.3.4.2 LED results: $\gamma = 0$

A preliminary LED simulation illustrates PointMap dynamics with information value computation based solely on the predictive accuracy $(\gamma = 0)$. Analysis of the deficiencies in the structure of the resulting code point, once again, to the importance of including criticality in the information value calculation.

In the simulation illustrated in Figure 6-7a, which shows the number of nodes stored each class *0* … *9*, PointMap created an unbalanced code, with many nodes assigned to some classes and few, or none, to others. This imbalance occurred because the randomly flipped line segments had a differential effect on different classes. For example, flipping one of the segments in the image **8** could change it into a noise-free

representative of class *0*, *6*, or *9*, though still labeled as belonging to class *8*. This property lowers the predictive accuracy, and hence the information value, of nodes that correctly encode *8*. In this example, where the information value is based entirely on predictive accuracy, class *8* was unable to retain any coding nodes. Similarly, one flip could change an image 3 into 9. Correspondingly class *3* retained only one of the 40 coding nodes, with that node storing a noisy image. On the other hand, flipping any one segment in an image 2 would produce a non-digit, which is still likely to be chosen as the nearest neighbor to a test-set exemplar of class *2*. Exemplars for class *2* thus produced high predictive accuracy values and six coding nodes, but, with the total code restricted in size, this overrepresentation of some classes contributed indirectly to the high test set error rate of others.

Figure 6-7b shows details of the PointMap code at the end of this preliminary simulation. Beneath the index of each coding node is the LED image of its first seven stored components, its assigned output class, and its final information value. Note that the images of only two of the forty stored nodes (#33 and #38) were noisy. This is much better than chance because the 10% segment flip probability implies that fewer than one of every two input images is noise-free. Figure 6-7b also shows that, except for the most recently added node (#40) retained nodes had similar information values.

With $\gamma = 0$ and $k = 1$, PointMap achieved classification accuracy of 49% on the test set. This result is better than the 41% achieved by the standard *1*-NN classifier, storing all training inputs, even though PointMap could not classify correctly any testing points from the class *8*. Nonetheless, this accuracy rate is far from the optimal 74%.

6.3.4.3 LED results: $\gamma = 0.25$

All PointMap simulations described so far have based predictions on the output class of single nearest neighbors $(k = 1)$. The irrelevant input components #8-24 in the LED example introduce the curse of dimensionality (Cybenko, Saarinen, Gray, Wu, & Khrabrarov, 1994). As the preliminary simulations in Section 3.4.2 have illustrated, PointMap with $k = 1$ does not produce satisfactory results under these circumstances. We will now see that higher values of k do produce near-optimal predictions. In these simulations, tie votes, which are rare, are broken in favor of the smallest output class number.

Crosses in Figure 6-8 represent results obtained by Wilson and Martinez on 16 types of *3*-NN algorithms. The solid line and circles in Figure 6-8 plot the average test-set accuracy achieved by PointMap for various values of the maximum code size C_{max}. Here, the number of nearest neighbors k was chosen by 10-fold cross-validation on the training set. The dashed line, which shows that a rule-of-thumb that simply sets $k = C_{max}/10$ produces near-optimal results, indicates the robustness of this parameter choice. Post-training pruning might further improve PointMap performance.

Note that Point Map selects test-set values of k which are high compared to typical values, which are usually less than 10 (Alpaydin, 1997). This result might be due to the fact that PointMap uses more than one nearest neighbor only during testing, so the

meaning of setting $k>1$ is not the same as in a standard k-NN algorithm, which normally uses the same k values during training and testing.

6.4 Summary and discussion

This study introduced PointMap, a simple memory-based learning system that computes the information value of the coding nodes to prune non-informative nodes during and after training. Two modes of operation of PointMap gave good results. Either the information value was based only on the predictive accuracy of each node, or it combined predictive accuracy and a criticality factor that forced the nodes to be closer to the decision boundary. The first choice is simpler, whereas the second one is more robust and it can give better performance if a good balance between the two factors is found. Comparison of PointMap to several code reduction techniques for the k-*NN* algorithm showed that PointMap was able to perform as accurately as any of them, with the additional advantage that the user can determine the resulting size of the created code.

There are several related studies available in the literature. First, PointMap shares certain features with the incremental pruning NN algorithms like CNN, IB2, IB3, and the Grow and Learn algorithm (Wilson and Martinez, 2000). However, the on-line pruning mechanism eliminates the main problem of these algorithms, their tendency to consider noisy data as significant exceptions that need to be included into code. In terms of computation of the *information value* PointMap keeps track of how many times a given node was the nearest neighbor, how many times it caused predictive error, etc. In this aspect it is similar to the MCS algorithm (Brodley, 1993). Finally, PoinMap has the ability to relearn in a non-stationary environment. There are several studies that look at systems in environments with this so-called concept shift (the optimal mapping changes over time) or sampling shift (data from a certain region of the input space presented together). For a review of methods used for this kind of data in connection with memory-based learning systems see Kuh, Petsche and Rivest (1991) or Salganicoff (1997). However, the present study does not evaluate PointMap's behavior with non-stationary data.

PointMap presents several directions for future development. Among them are: evaluation of the choice of the end condition on the system behavior, automatic determination of the system parameters, the possibility of using different values of the criticality parameter γ for each node to better approximate variable-complexity decision boundaries, introduction of temporal degradation of information value to eliminate rarely activated nodes, exploration of alternative methods of on-line pruning, and evaluation of PointMap behavior on non-stationary data. On the other hand, the methods of information-value-based on-line and post-training pruning introduced here can be readily applied in more complex memory-based learning systems like fuzzy ARTMAP (Carpenter et al., 1992) and the Nested Generalized Exemplar (Salzberg, 1990).

Chapter 7 Graded Signal Functions for ARTMAP Neural Networks[2]

Abstract

This study presents an analysis of a modified ARTMAP neural network in which a graded signal function replaces the standard choice-by-difference function. The modifications are introduced mathematically and the performance of the system is studied on two benchmark examples. It is shown that the modified ARTMAP system achieves classification accuracy superior to that of standard ARTMAP, while retaining comparable complexity of the internal code.

7.1 Introduction

The present chapter focuses on the process of search for a best category code in response to a given input in ARTMAP networks. Specifically, a new signal function is proposed that enables the system to find near-optimum discrimination curves between categories in complex input space. The modified signal function is introduced mathematically, then evaluated by implementing it in a fuzzy ARTMAP system and analyzing its performance on two benchmark problems.

7.2 Description of Fuzzy ARTMAP Dynamics

This section gives a brief summary of the fuzzy ARTMAP algorithm (Carpenter, Grossberg, Markuzon, Reynolds, and Rosen, 1992). The **inputs** of the fuzzy ARTMAP system are usually normalized by complement coding, which converts an M-dimensional input vector $\boldsymbol{a} = (a_1,\ldots,a_M)$ $(0 \le a_i \le 1)$ into $2M$-dimensional input pattern $\boldsymbol{I} = (\boldsymbol{a}, 1-\boldsymbol{a}) = (a_1,...,a_M,1-a_1,...,1-a_M)$.

The pattern is normalized since $|\boldsymbol{I}| = M$, where $|\boldsymbol{I}| \equiv \sum_{i=1}^{2M} |I_i|$ is the city-block norm.

When a new input is presented, the system **searches** for a candidate coding node within its coding layer. In ART systems, the j-th coding node defines a hyper-rectangle R_j, or *coding box*, in the M-dimensional input space, described by the weights $w_{i,j}$ leading to that node. The hyper-rectangle reduces to a rectangle in two dimensions or to an interval in one dimension (Figure 7-1). For every input pattern, the ARTMAP search mechanism chooses the smallest coding box that is covering the input, or the box that is closest to the input, based on the activation of the *choice-by-difference* (CBD) signal function T_j (Carpenter and Gjaja, 1994), now used in a majority of simulations. The CBD signal function is defined as:

$$T_j = M(2-\alpha) - d(R_j,\boldsymbol{a}) - \alpha|R_j| = |\boldsymbol{w}_j \wedge \boldsymbol{I}| + (1-\alpha)\left(2M - |\boldsymbol{w}_j|\right). \qquad (1)$$

In this equation, α is a parameter (usually $\alpha=0^+$), $d(R_j,\boldsymbol{a})$ represents the city-block distance from the input pattern $\boldsymbol{a}$ to the coding box R_j, and $|R_j|$ represents the size of R_j. The J-th coding node is chosen as a candidate code if its signal function T_J has the maximum value.

---

[2] Published in Sinčák et al. (Eds.) The State of the Art in Computational Intelligence (Collection of papers presented at the Symposium on Computational Intelligence, Košice, Slovakia, 2000).

The candidate node is then compared with the input pattern according to a **match rule**. The candidate **resonates** if $|\boldsymbol{I} \wedge \boldsymbol{w}_J| > \rho \boldsymbol{I}$; or it is **reset** if the inequality does not hold, where $\rho \in [0,1]$ is called a *vigilance* parameter. If reset occurs, a search for a new candidate is initiated, or a new coding node is created. If the candidate node resonates, the system checks whether the node is associated with the correct output class (always satisfied for new nodes). If the node is associated with an incorrect category, a process of **match tracking** is initiated, i.e., ρ is increased just enough so that the current candidate will not resonate any more and a search for a new candidate is initiated.

Once a coding node is found that satisfies all the requirements, **learning** is initiated that updates all the weights leading to the J-th node, defined by

$$\boldsymbol{w}_J^{(new)} = \beta\left(\boldsymbol{I} \wedge \boldsymbol{w}_J^{(old)}\right) + (1-\beta)\boldsymbol{w}_J^{(old)}. \tag{2}$$

Fast learning is usually chosen, obtained by setting $\beta=1$.

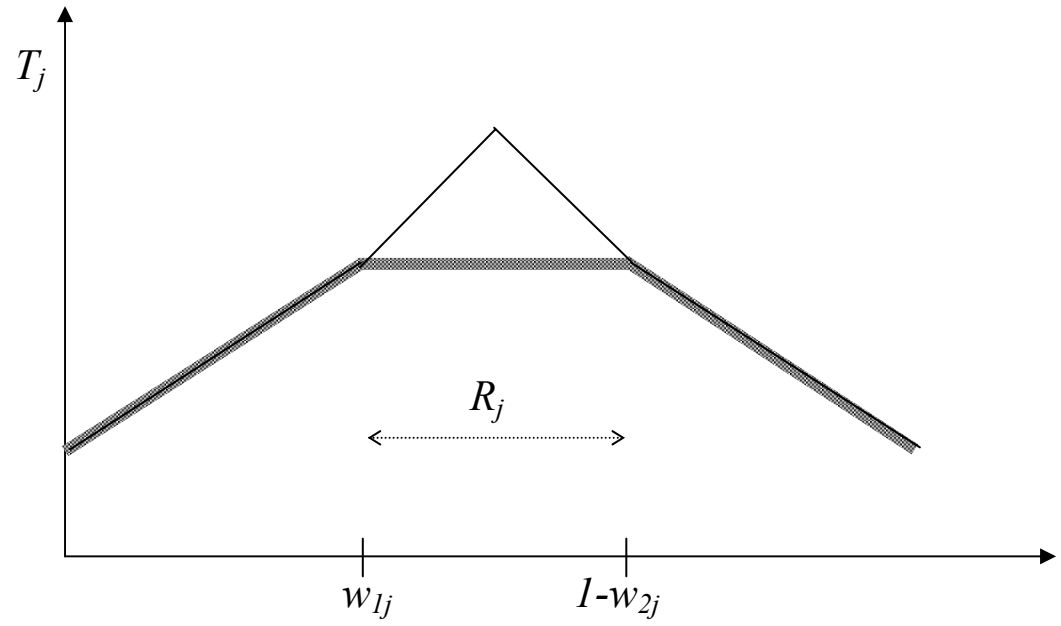


Figure 7-1 Choice signal for standard CBD (▬) vs. graded CBD (—) in one input dimension

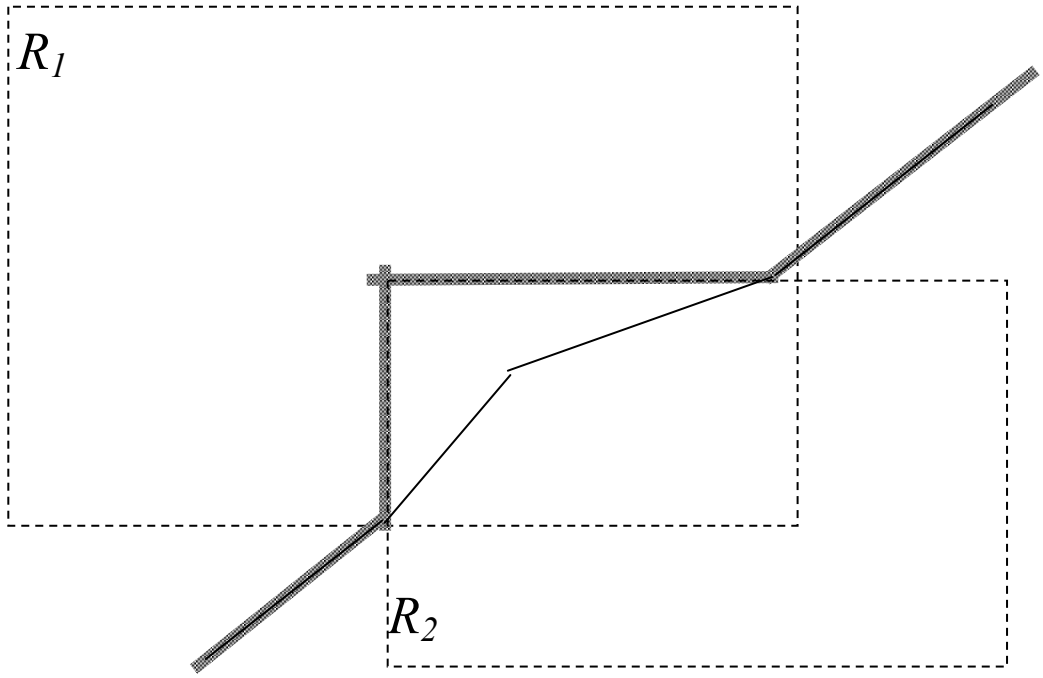


Figure 7-2 Decision boundaries between two category boxes (R_1 and R_2) with standard CBD (▒) vs. graded CBD (—)

7.3 Definition of Graded Signal Functions

In general, the ARTMAP search for the internal code that best matches the presented input pattern can be accomplished by choosing one of many different signal functions, used to determine the activation of the coding nodes. Most current ARTMAP systems use the *choice-by-difference* signal function (CBD), which implements the idea of minimum fast learning. For the CBD function the signal is independent of the position of the input pattern if the input is located within the coding box, as shown for one dimension in Figure 7-1. In the present chapter, a new signal function is introduced, called a *graded choice-by-difference* function, or graded CBD, which makes the choice signal dependent on the input position even when the input lies within the category box. Namely, an input near the center of the box R_j generates a larger signal T_j than an input near the boundary of the box (Figure 7-1).

The activation in the graded CBD signal function is defined by:

$$T_j = M(2-\alpha) - d(R_j, \boldsymbol{a}) - \alpha\left|R_j\right|(1-\eta\gamma_j) \tag{3}$$

where η is a parameter that defines by how much the activation at the center of the box is increased relative to the box boundaries. When η=0, graded CBD reduces to standard CBD (1). In (3), γ_j specifies the minimum of graded activations across dimensions *i=1...M*:

$$\gamma_j = \min_{i=1..M}\left[1-\left[\frac{2\left|a_i - c_{j,i}\right|}{1-w_{j,i+M}-w_{j,i}}\right]^+\right]^+ \tag{4}$$

In (4), $c_{j,i} \equiv (1-w_{j,i+M}+w_{j,i})/2$ denotes the center of the *j*-th coding box in the *i*-th dimension, and $[a]^+ \equiv \max(a,0)$ is a rectification operator. Note that $\gamma_j = 1$ at the center of R_j and $\gamma_j = 0$ at any point $\boldsymbol{a}$ on the boundary of R_j. In order to ensure that the same input would choose the same category if it were immediately re-presented (direct access), the ART match rule was also modified, to better correspond to the new choice rule. In addition to the match criterion defined above, the new match rule essentially simulates the process of weights-update (2) followed by re-presentation of the current input. Then, the *J*-th node resonates only if the simulated update led to the desired choice of the winning node by direct access. The resulting graded signal function system has the capacity to create more accurate decision boundaries, especially when these boundaries are not parallel to the input space axes (Figure 7-2).

7.4 Results on Benchmark Data and Discussion

The performance of a fuzzy ARTMAP system with the graded CBD signal rule was evaluated on two benchmark problems, the circle-in-the-square (CIS) problem and a diagonal problem. Data sets for both problems consist of 2-dimensional uniformly distributed points, with the values in each dimension ranging from 0 to 1. Each data set has two output classes. In CIS, a point $\boldsymbol{a}=(a_1,a_2)$ is in the class C_{out} if $(a_1-0.5)^2+(a_2-0.5)^2 > \frac{1}{2\pi}$, otherwise it is from the class C_{in}. In the diagonal data set a point is from the class C_{lower} if $a_1>a_2$, otherwise it is from the class C_{upper}. Simulations with training sets of different sizes (100, 1000, or 10,000 points) were performed. The testing set size was fixed to 10,000 points.

The results of simulations of the two benchmark problems are shown in Figure 7-3, which shows percent correct predictions and number of coding nodes as functions of the value of the graded signal parameter η. In the graphs, η=0 corresponds to the standard CBD rule. Each point in Figure 7-3 corresponds to an average of 10 simulations with randomized order of inputs. For each of the conditions, application of the graded signal function led to improved performance, accompanied in some conditions by a slight increase in the internal code complexity. This improvement is mainly due to the improved ability of the new signal function to approximate decision boundaries not parallel to the axes of input feature space (Figure 7-2).

These results indicate that ARTMAP systems with the graded CBD signal rule can be used for many types of pattern recognition problems, especially when the data

from individual classes are not easily separable, which may lead to many overlapping category boxes. More simulations are necessary, especially with noisy data, to better understand the behavior of the system in complex environments.

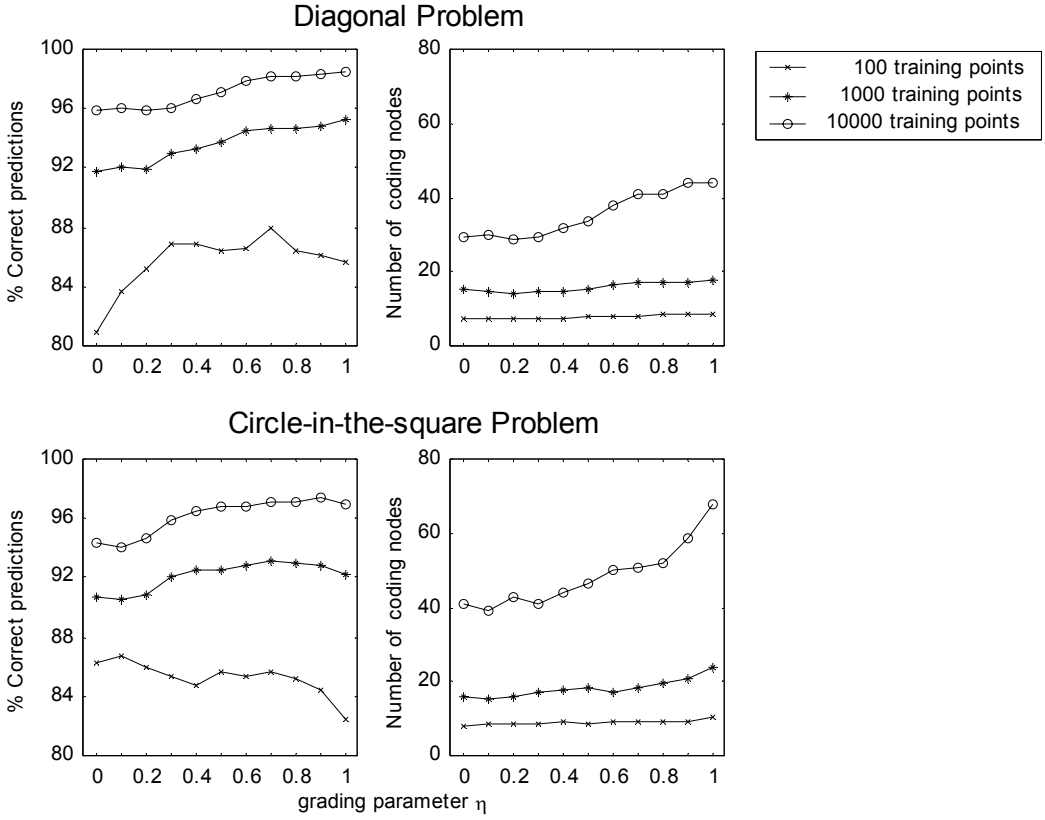


Figure 7-3 Simulations of fuzzy ARTMAP with standard CBD (η=0) and with graded CBD signal function (η>0). The upper row shows results of simulations with the diagonal data set, the lower row contains data for circle-in-the-square simulations

Chapter 8 Summary, conclusions, and directions for future work

The results described in this dissertation contribute to our understanding of how human listeners and computational learning algorithms cope with complex, noisy environments. The psychoacoustic studies present new insights into how humans detect masked sounds in anechoic rooms and how they localize nearby sound sources in reverberant environments. Computational learning algorithms are proposed that use the methods of memory-based learning to effectively encode complex noisy data. The following sections summarize the results of these studies and propose future work on each of the studied topics.

8.1 Detection of pure-tone sources masked by noise

Spatial unmasking of pure tone stimuli in a simulated anechoic environment was measured and modeled. This work bridges the gap between past headphone studies of binaural unmasking and past free-field studies of spatial unmasking. In addition, to our knowledge this is the first study to look at spatial unmasking of nearby sources and to examine spatial unmasking as a function of source distance. Results show:

1. Spatial unmasking is large for nearby sources, mainly due to large changes in the received level associated with changes in spatial position of the sources.
2. In general, azimuthal separation of target and masker leads to unmasking. However, when the masker is laterally displaced relative to the listener, there are some conditions for which the amount of masking either remains constant or even increases slightly when the target is displaced from the masker (towards the median plane).
3. The energetic and binaural cues are approximately equally important for spatial unmasking at lower frequencies (500 Hz), while the energetic cues dominate at higher frequencies (1 kHz).
4. No condition was found where monaural better ear performance is better than binaural performance.
5. Model predictions of spatial masking that take into account both acoustic changes at the ears and binaural interaction capture all important trends in across-subject average performance. However, current models cannot account for individual subject differences in the contributions of binaural processing and how this contribution varies with target and masker positions.

These data suggest that the distribution of binaural coincidence detectors might differ from subject to subject, an idea that needs to be addressed with further research. Further work is also needed in order to understand and predict spatial unmasking for more complex stimuli like click-trains, tone complexes, and speech. While some data examining this question exists, models that predict spatial unmasking of non-speech complex stimuli are virtually non-existent. Finally, in order to relate these results to more natural, real-world settings, future experiments investigating the effect of reverberation on spatial unmasking are necessary.

8.2 Localization in reverberant rooms

A study of the effect of experience and listener position on localization performance was performed with listeners at four different positions in an ordinary classroom. Two listener groups were used to control for the effect of experience. Previous acoustic analyses of the effects of the room on received sounds for the four listener locations allowed direct comparison of acoustics and behavior. Results suggest that:

1. Both room position and experience have a measurable impact on localization performance.
2. The variability in subject responses (in both azimuth and distance) was influenced by both room acoustics and listener experience.
    a. Response variability decreased with the acoustic complexity of the listener location.
    b. Experience on the task and in the room decreased response variability.
3. Response bias (mean signed localization error) was not consistently affected by either acoustics or experience:
    a. There was no consistent effect on mean perceived distance.
    b. When the listener was located with the wall to his back, there was a small but consistent azimuthal bias.
4. The effect of room position was stronger for nearby sources.
5. The effect of experience was greater for far sources.
6. No learning was observable within a single session.
7. Predictions based on acoustic analysis did not match the behavioral data very well.

The main questions that remain to be answered by future studies are:

1. Is there any consistent effect of experience and room position on bias in azimuthal and distance perception?
2. How does experience influence localization performance?
    a. What is the time course of performance changes?
    b. Is the effect of experience influenced by listener position?
    c. What gets learned as the listener gets more experienced with the room?

Both within-subject and inter-subject variability were large in the current experiment, making it hard to draw firm conclusions. A possible solution to this problem is to perform follow-up studies in a simulated environment where the acoustic cues are perfectly controlled and noise in the response-measuring equipment is minimized. In order to model these behavioral results, simple acoustic measurements should be used to provide realistic inputs to peripheral models of auditory processing in order to obtain predictions that are more closely based on the actual processing in the brain.

8.3 Pattern recognition

The following two studies introduced new methods for improvement of classification performance of two learning algorithms when exposed to noisy data.

The first study introduced PointMap, a simple memory-based learning system that computes information value of the coding nodes to prune non-informative nodes during and after training. Performance of PointMap was found to be very good compared to multiple other memory-based learning systems. Future development of the PointMap system can include:

1. Evaluation of the choice of the end condition on the system behavior.
2. Automatic determination of the system parameters.
3. Exploring the possibility of using different value of the γ parameter for each node to better approximate variable-complexity decision boundaries.
4. Introduction of temporal degradation of information value to eliminate rarely activated nodes.
5. Exploration of alternative methods of on-line pruning, and evaluation of PointMap behavior on non-stationary data.
6. Applicability of information-value-based pruning methods for more complex memory-based learning systems like fuzzy ARTMAP and the Nested Generalized Exemplar.

Of course, the most important follow-up on development of any new learning algorithm is to extensively test it in various applications.

The final study presented in this thesis was introduction of graded signal functions into the fuzzy ARTMAP algorithm. The most recent development in the ART family of neural networks is the distributed ARTMAP algorithm. The graded signal function can be used also in this ART algorithm. However, the main prerequisite for its applicability is to develop a suitable match function that can secure balance in the performance of the ART algorithms with the new signal function.

Chapter 9 Appendix

9.1 Independence of responses in study of spatial unmasking

The behavioral experiment performed as part of the study of spatial unmasking (Chapter 3) used an adaptive procedure (3-down-1-up, three-interval two-alternative forced-choice procedure) to obtain the detection thresholds (Levitt, 1971). The procedure tracked the 79.4%. During the experiment the subjects appeared to react differently depending on the feedback. Namely, incorrect response at any level seemed to make it more probable that the subject will respond incorrectly in the following trial. Such a behavior could have two possible reasons. First, subjects could be changing their concentration on the task depending on the feedback. For example, the feedback saying that the previous response was incorrect could increase the subject's concentration on the following trial, whereas positive feedback would lead to no modification of his/her attentional level. Second, it is also possible that the subjects learn spontaneously the adaptive procedure, and change their behavior accordingly. This appendix presents the analysis of significance of this effect, and looks at possible influence it could have on the measured thresholds.

The t-test of significance was used to determine whether any such effect is observable. First, an estimate of the psychometric function was computed for every spatial configuration and for all the levels on the psychometric function for which at least two measurements were taken. An example of this computation is shown for one subject and one spatial configuration is shown in Figure 9-1.

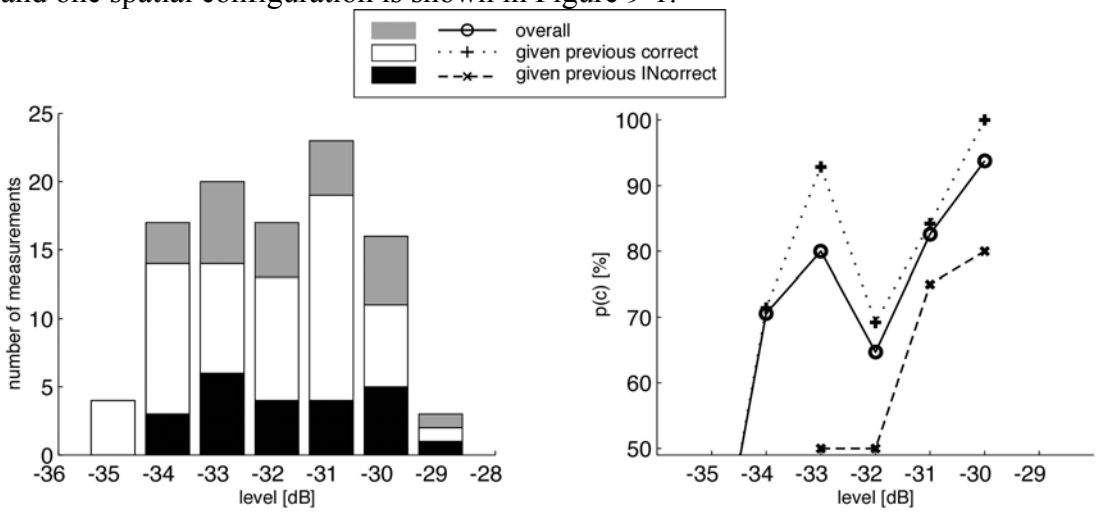


Figure 9-1 Example of the procedure used for estimation of the psychometric functions. The left graph shows, for one subject and one spatial condition, the number of measurements at different levels, given that the previous response was correct, given that previous response was incorrect, and overall. The data were divided into these three groups and the psychometric function estimates (the right-hand graphs) were computed for the corresponding functions.

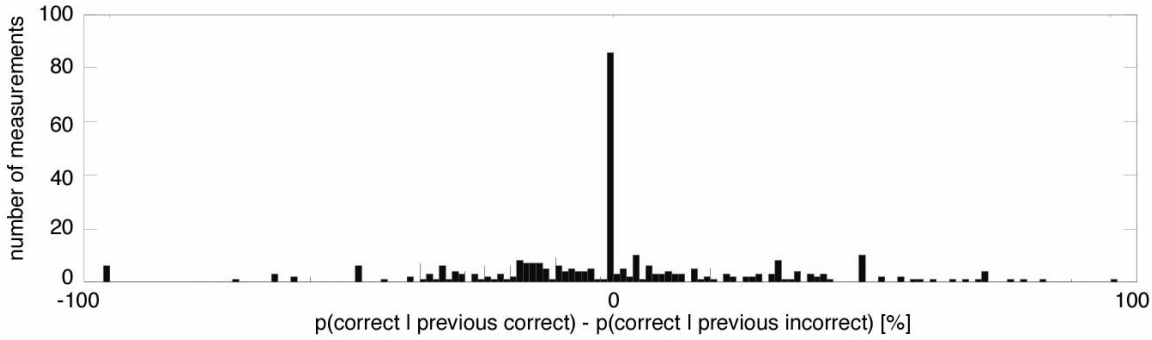


Figure 9-2 Example of the distribution of differences in the estimates of points on the psychometric function, given previous response was correct vs. previous response incorrect. Data for one subject, collapsed across spatial configurations and presentation levels.

Then, a difference in the psychometric function was computed for every spatial configuration and every presentation level. The difference values were then collapsed across spatial configurations and presentation levels, resulting in distributions similar to the one in Figure 9-2.

The t-test tested the hypothesis that there is no significant difference between the two means, i.e.,

H_0: E{ p(correct | previous correct) - p(correct | previous not correct)} = 0

The results are summarized in

Table 9-1. When looking at individual subject performance at both frequencies (upper six rows), the effect is significant for subject S1 at both frequencies. The most significant effect is observed for S3 at 1000 Hz, but this time the shift in the mean values between the two conditions is in the opposite direction, i.e., the subject was more probable to respond correctly if the previous response was correct than when it was incorrect. Collapsing data across subjects led to a significant effect at 500 Hz, but not at 1000 Hz. This is to be expected because at 1000 Hz the performance is shifted in opposite directions for subjects S1 and S3, and it is essentially unbiased for subject S2. Collapsing across subjects and frequencies led to a marginally significant effect (p=0.084) in the expected direction. Because this result is indecisive, a simulation was performed to estimate the influence this conditional dependence of response on previous response could have on the measured thresholds. A condition was simulated corresponding to the performance of the subject S3 in the 1000 Hz condition, where the difference in the means was the largest. In the simulations it was assumed that the subject's response in the two conditions (given previous response correct and incorrect) can be approximated by a Gaussian distributions with the means separated by 4.74 % and

standard deviation of 10.1, estimated from subject S3's performance in the 1000 Hz condition. The result of a simulation with 10,000 trials is shown in Figure 9-3. This figure shows three psychometric functions, one obtained when looking only at responses preceded by a correct response, one for responses preceded by incorrect responses, and the overall psychometric function including both conditions. The figure shows that the difference between thresholds estimated from responses preceded by correct responses and those preceded by incorrect responses is less than 0.5 dB. Therefore it can be concluded that, although there can be significant shifts in the thresholds depending on response in the previous trials, these differences are relatively small compared to the changes in the estimated detection threshold due to other factors.

Table 9-1 Results of the t-test of significance on the dependence of subject's response on feedback from previous trial.

	N	M_1-M_2	t_{obs}	p	
500 Hz S1	405	3.66	2.185	0.015	REJECT
500 Hz S2	383	0.86	0.533	0.297	
500 Hz S3	370	0.80	0.480	0.316	
1000 Hz S1	407	3.88	2.259	0.012	REJECT
1000 Hz S2	388	0.17	0.108	0.457	
1000 Hz S3	343	-4.74	-3.098	0.001	REJECT
500 Hz x-S	1158	1.82	1.907	0.028	REJECT
1000 Hz x-S	1138	0.02	0.018	0.493	
x-freq x-subj	2296	0.93	1.381	0.084	

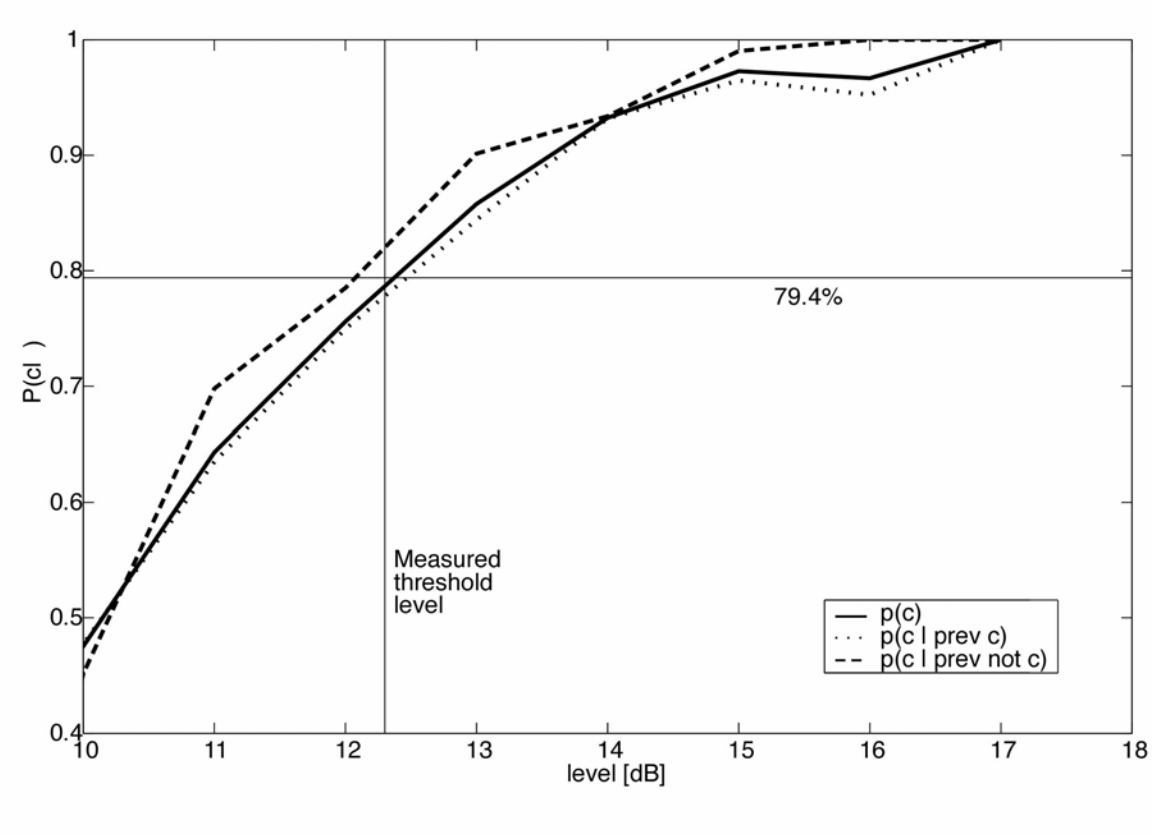


Figure 9-3 Psychometric functions estimated in a simulation assuming that subjects' behavior is dependent on feedback and previous response.

REFERENCES

Aha DW (1997) Lazy learning. Artificial Intelligence Review 11:7-10.

Aha DW, Kibler D, Albert MK (1991) Instance-based learning algorithms. Machine Learning 6:37-66.

Alpaydin E (1997) Voting over multiple condensed nearest neighbors. Artificial Intelligence Review 11:115-132.

Blauert J (1997) Spatial Hearing, 2nd Edition. Cambridge, MA: MIT Press.

Brodley C (1993) Addressing the selective superiority problem: Automatic algorithm/model class selection. In: 10th Intl. Conf. Machine Learning, pp 17-24: Morgan Kaufmann.

Bronkhorst AW (2000) The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions. Acustica 86:117-128.

Bronkhorst AW, Plomp R (1988) The effect of head-induced interaural time and level differences on speech intelligibility in noise. Journal of the Acoustical Society of America 83:1508-1516.

Bronkhorst AW, Houtgast T (1999) Auditory distance perception in rooms. Nature 397:517-520.

Brown TJ (2001) 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.

Brungart DS (1998) Near-field auditory localization. Unpublished Ph. D. Thesis. Cambridge, MA: Massachusetts Institute of Technology. Electrical Engineering and Computer Science Department.

Brungart DS, Rabinowitz WM (1999) Auditory localization of nearby sources I: Head-related transfer functions. Journal of the Acoustical Society of America 106:1465-1479.

Brungart DS, Durlach NI (1999) Auditory localization of nearby sources II: Localization of a broadband source in the near field. Journal of the Acoustical Society of America 106:1956-1968.

Carpenter GA (1997) Distributed learning, recognition, and prediction by ART and ARTMAP neural networks. Neural Networks 10:1473-1494.

Carpenter GA, Grossberg S (1987a) A massively parallel architecture for a self-organizing neural pattern recognition machine. Computer Vision, Graphics, and Image Processing 37:54-115.

Carpenter GA, Grossberg S (1987b) ART2: self-organization of stable category recognition codes for analog input patterns. Applied Optics 26:4919-4930.

Carpenter GA, Grossberg S, eds (1991) Pattern Recognition by Self-Organizaing Neural Networks. Cambridge, MA: Bradford Books, M. I. T. Press.

Carpenter GA, Grossberg S, Rosen DB (1991) Fuzzy ART: fast stable learning and categorization of analog patterns by an adaptive resonance system. Neural Networks 4:759-771.

Carpenter GA, Grossberg S, Reynolds JH (1991) ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network. Neural Networks 4:565-588.

Carpenter GA, Milenova BL, Noeske BW (1998) Distributed ARTMAP: a neural network for fast distributed supervised learning. Neural Networks 11:793-813.

Carpenter GA, Grossberg S, Markuzon N, Reynolds JH, Rosen DB (1992) Fuzzy ARTMAP: a neural network architecture for incremental supervised learning of analog multidimensional maps. IEEE Transactions on Neural Networks 3:698-713.

Cherry EC (1953) Some experiments on the recognition of speech, with one and with two ears. Journal of the Acoustical Society of America 25:975-979.

Colburn HS (1977a) Theory of binaural interaction based on auditory-nerve data. II: Detection of tones in noise. Journal of the Acoustical Society of America 64:525-533.

Colburn HS (1977b) Theory of binaural interaction based on auditory-nerve data. II: Detection of tones in noise. Journal of the Acoustical Society of America supplementary material, AIP document no. PAPS JASMA-61-525-98.

Colburn HS (1996) Binaural models. In: Auditory Computation (Hawkins HL, McMullen TA, Popper AN, Fay RR, eds), pp 332-400. New York: Springer Verlag.

Colburn HS, Durlach NI (1978) Models of binaural interaction. In: Handbook of perception (Carterette EC, Friedman MP, eds), pp 467-518. New York: Academic Press.

Coleman PD (1968) Dual role of frequency spectrum in determination of auditory distance. Journal of the Acoustical Society of America 44:631-632.

Cover TM, Hart PE (1967) Nearest neighbor pattern classification. IEEE Transactions on Information Theory 13:21-27.

Cybenko G, Saarinen S, Gray R, Wu Y, Khrabrov A (1994) On the effectiveness of memory-based methods in machine learning. In: Dealing with complexity: A neural networks approach (Karny M, Warwick K, Kurkova V, Taylor JG, eds), pp 62-75. New York: Springer Verlag.

Dasarathy BV (1991) Nearest Neighbour (NN) Norms: NN Pattern Classification Techniques. Los Alamitos, CA: IEEE Computer Society Press.

Dasarathy BV, Sánchez JS, Townsend S (2000) Nearest neighbour editing and condensing tools - synergy exploitation. Pattern Analysis and Applications 3:19-30.

Doll TJ, Hanna TE (1995) Spatial and spectral release from masking in threedimensional auditory displays. Human Factors 37:341-355.

Doll TJ, Hanna TE, Russotti JS (1992) Masking in threedimensional auditory displays. Human Factors 34:255-265.

Duda RO (1997) Elevation dependence of the interaural transfer function. In: Binaural and Spatial Hearing in Real and Virtual Environments (Gilkey R, Anderson T, eds), pp 49-76. New York: Erlbaum.

Duda RO, Martens WL (1998) Range dependence of the response of a spherical head model. Journal of the Acoustical Society of America 104:3048-3058.

Duda RO, Hart PE, Stork DG (2001) Pattern Classification, 2nd Edition: Wiley-Interscience.

Durlach NI (1972) Binaural signal detection: Equalization and cancellation theory. In: Foundations of Modern Auditory Theory (Tobias JV, ed). New York: Academic Press.

Durlach NI, Colburn HS (1978) Binaural phenomena. In: Handbook of Perception (Carterette EC, Friedman MP, eds), pp 365-466. New York: Academic Press.

Ebata M, Sone T, Nimura T (1968) Improvement of hearing ability by directional information. Journal of the Acoustical Society of America 43:289-297.

Freyman RL, Balakrishnan U, Helfer K (2000) Release from informational masking in speech recognition. In: MidWinter Meeting of the Association for Research in Otolaryngology, p 89. St. Petersberg Beach, FL.

Freyman RL, Helfer KS, McCall DD, Clifton RK (1999) The role of perceived spatial separation in the unmasking of speech. Journal of the Acoustical Society of America 106:3578-3588.

Gatehouse RW (1987) Further research on free-field masking. Journal of the Acoustical Society of America 82:S108.

Gilkey R, Anderson T (1997) Binaural and Spatial Hearing in Real and Virtual Environments. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc.

Good MD, Gilkey RH, Ball JM (1997) The relation between detection in noise and localization in noise in the free field. In: Binaural and Spatial Hearing in Real and Virtual Environments (Gilkey R, Anderson T, eds), pp 349-376. New York: Erlbaum.

Grossberg S (1976) Adaptive pattern classification and universal recoding. II: Feedback, expectation, olfaction, and illusions. Biological Cybernetics 23:187-202.

Hart PE (1968) The condensed nearest neighbor rule. IEEE Transactions on Information Theory 14:515-516.

Hartmann WM (1983) Localization of sound in rooms. Journal of the Acoustical Society of America 74:1380-1391.

Hartmann WM (1997) Listening in a room and the precedence effect. In: Binaural and Spatial Hearing in Real and Virtual Environments (Gilkey R, Anderson T, eds), pp 191-210. New York: Erlbaum.

Hawley ML, Litovsky RY, Colburn HS (1999) Speech intelligibility and localization in a multi-source environment. Journal of the Acoustical Society of America 105:3436-3448.

Kidd G, Mason CR, Rohtla TL, Deliwala PS (1998) Release from masking due to spatial separation of sources in the identification of nonspeech auditory patterns. Journal of the Acoustical Society of America 104:422-431.

Kidd G, Mason CR, Deliwala PS, Woods WS, Colburn HS (1994) Reducing informational masking by sound segregation. Journal of the Acoustical Society of America 95:3475-3480.

Kuh A, Petsche T, Rivest RL (1991) Learning time-varying concepts. In: Advances in Neural Information Processing Systems, pp 183-189: Morgan Kaufmann.

Lam W, Keung C-K, Liu D (2002) Discovering useful concept prototypes for classification based on filtering and abstraction. IEEE Transactions on Pattern Analysis and Machine Intelligence 24:1075-1090.

Levitt H (1971) Transformed up-down methods in psychophysics. Journal of the Acoustical Society of America 49:467-477.

Lutfi RA (1990) How much masking is informational masking? Journal of the Acoustical Society of America 88:2607-2610.

Makous JC, Middlebrooks JC (1990) Two-dimensional sound localization by human listeners. Journal of the Acoustical Society of America 87:2188-2200.

Merz CJ, & Murphy, P. M. (1996) UCI repository of machine learning databases. In.: Irvine, CA: University of California Irvine, Department of Information and Computer Science URL: http://www.ics.uci.edu/~mlearn/MLRepository.html.

Middlebrooks JC, Green DM (1991) Sound localization by human listeners. Annual Review of Psychology 42:135-159.

Moore BCJ (1997) An Introduction to the Psychology of Hearing (4e). San Diego, CA: Academic Press.

Musicant AD, Butler RA (1985) Influence of monaural spectral cues on binaural localization. Journal of the Acoustical Society of America 77:202-208.

Rakerd B, Hartmann WM (1985) Localization of sound in rooms. II. The effects of a single reflecting surface. Journal of the Acoustical Society of America 78:524-533.

Rakerd B, Hartmann WM (1986) Localization of sound in rooms. III. Onset and duration effects. Journal of the Acoustical Society of America 80:1695-1706.

Ram A (1993) Indexing, elaboration, and refinement: Incremental learning of explanatory cases. Machine Learning 10:201-248.

Saberi K, Dostal L, Sadralodabai T, Bull V, Perrott DR (1991) Free-field release from masking. Journal of the Acoustical Society of America 90:1355-1370.

Salganicoff M (1997) Tolerating concept and sampling shift in lazy learning using prediction error context switching. Artificial Intelligence Review 11:133-155.

Salzberg SL (1990) Learning with Nested Generalized Exemplars. Hingham, MA: Kluwer Academic.

Santarelli S (2000) Auditory localization of nearby sources in anechoic and reverberant environments. Unpublished Ph. D. Thesis. Boston, MA: Boston University. Cognitive and Neural Systems Department.

Santarelli S, Kopčo N, Shinn-Cunningham BG (1999a) Localization of near-field sources in a reverberant room. In: 22nd mid-Winter meeting of the Association for Research in Otolaryngology, p 26. St. Petersburg Beach, FL.

Santarelli S, Kopčo N, Shinn-Cunningham BG, Brungart DS (1999b) Near-field localization in echoic rooms. Journal of the Acoustical Society of America 105:1024.

Santon F (1987) Détection d'un son pur dans un bruit masquant suivant l'angle d'incidence du bruit. Relation avec le seuil de réception de la parole (Detection of a pure sound in the presence of masking noise, and its dependence on the angle of

incidence of noise. Relation with the speech reception threshold). Acustica 63:222-228.

Shinn-Cunningham BG (2000) Learning reverberation: Considerations for spatial auditory displays. In: Proceedings of the International Conference on Auditory Displays, pp 126-134. Atlanta, GA.

Shinn-Cunningham BG, Santarelli S, Kopčo N (2000) Tori of confusion: Binaural localization cues for sources within reach of a listener. Journal of the Acoustical Society of America 107:1627-1636.

Shinn-Cunningham BG, Schickler J, Kopčo N, Litovsky RY (2001) Spatial unmasking of nearby speech sources in a simulated anechoic environment. Journal of the Acoustical Society of America 110:1118-1129.

Stanfill C, Waltz D (1986) Toward memory-based reasoning. Communications of the ACM 29:1213-1228.

Stern RM, Shear GD (1996) Lateralization and detection of low-frequency binaural stimuli: Effects of distribution of internal delay. Journal of the Acoustical Society of America 100:2278-2288.

Strutt JW (1907) On our perception of sound direction. Philosophical Magazine 13:214-232.

Sycara KP, Navinchandra D (1989) A process model of experience-based design. In: Eleventh Annual Conference of the Cognitive Science Society, pp 283-290. Ann Arbor, MI: Lawrence Earlbaum.

van de Par S, Kohlrausch A (1999) Dependence of binaural masking level differences on center frequency, masker bandwidth, and interaural parameters. Journal of the Acoustical Society of America 106:1940-1947.

Vanderkooy J (1994) Aspects of MLS measuring systems. Journal of the Audio Engineering Society 42:219-231.

Wagenaars WM (1990) Localization of sound in a room with reflecting walls. Journal of the Audio Engineering Society 38:89-110.

Watson CS, Kelly WJ, Wroton HW (1976) Factors in the discrimination of tonal patterns II: Selective attention and learning under various levels of stimulus uncertainity. Journal of the Acoustical Society of America 60:1176-1186.

Wenzel EM, Arruda M, Kistler DJ, Wightman FL (1993) Localization using nonindividualized head-related transfer functions. Journal of the Acoustical Society of America 94:111-123.

Wettschereck D, Aha DW, Mohiri T (1997) A review and empirical evaluation of feature weighing methods for a class of lazy learning algorithms. Artificial Intelligence Review 11:273-314.

Wightman FL, Kistler DJ (1989) Headphone simulation of free-field listening. II. Psychophysical validation. Journal of the Acoustical Society of America 85:868-878.

Wightman FL, Kistler DJ (1997) Factors affecting the relative salience of sound localization cues. In: Binaural and Spatial Hearing in Real and Virtual Environments (Gilkey R, Anderson T, eds), pp 1-24. New York: Erlbaum.

Wilson DL (1972) Asymptotic properties of nearest neighbor rules using edited data. IEEE Transactions on Systems, Man, and Cybernetics 2:431-433.

Wilson DR, Martinez TR (1997) Improved heterogenous distance functions. Journal of Artificial Intelligence Research 6:1-34.

Wilson DR, Martinez TR (2000) Reduction techniques for instance-based learning algorithms. Machine Learning 38:257-286.

Zahorik P (2000) Loudness constancy with varying sound source distance. Nature Neuroscience 4:78-83.

Zurek PM (1993) Binaural advantages and directional effects in speech intelligibility. In: Acoustical Factors Affecting Hearing Aid Performance (Studebaker G, Hochberg I, eds). Boston, MA: College-Hill Press.

CURRICULUM VITAE

NORBERT KOPČO
kopco@bu.edu

Education:

1996 Technická Univerzita Košice, Slovakia, Ing. (M.Sc.) Electrical Engineering and Computer Science

Professional Experience:

1996 – 1997 Teaching Assistant – Department of Cybernetics and Artificial Intelligence, Technická Univerzita Košice

1998 – 2002 Research/Teaching Assistant – Cognitive and Neural Systems department, Boston University

Major Publications:

Shinn-Cunningham, BG, SG Santarelli, and N Kopčo (2000) "Tori of confusion: Binaural cues for sources within reach of a listener," J Acoust Soc Am, 107(3), 1627-1636.

Shinn-Cunningham, BG, J Schickler, N Kopčo, and RY Litovsky (2001). "Spatial Unmasking of Nearby Speech Sources in a Simulated Anechoic Environment", J. Acoust. Soc. Am, 110(2), 1118-1129.

Conference Papers:

Kopčo, N and GA Carpenter (2000) "Graded Signal Functions for ARTMAP Neural Networks," In Sinčák et al. (Eds.) The State of the Art in Computational Intelligence (Collection of papers presented at the European Symposium on Computational Intelligence, Košice, Slovakia, Aug 30 - Sept 1, 2000). Physica-Verlag. pp. 9-14.

Shinn-Cunningham, BG, JG Desloges, and N Kopčo (2001). "Empirical and modeled acoustic transfer functions in a simple room: Effects of distance and direction," in Proceedings of the 2001 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Pfaltz, NY, 19-24 October 2001, 183-186.

Kopčo, N and BG Shinn-Cunningham (2002) "Effects of Cuing on Perceived Location of Auditory Sources," In Sinčák et al. (Eds.) Intelligent Technologies - Theory and Applications (Collection of papers presented at the 2nd Euro-International Symposium on Computational Intelligence, Košice, Slovakia, June 16-19, 2002). IOS Press. pp. 201-209.

Kopčo, N and BG Shinn-Cunningham (2002) "Auditory Localization in Rooms: Acoustic Analysis and Behavior," to appear in the Proceedings of The 32nd International Acoustical Conference - EAA symposium "ACOUSTICS BANSKA STIAVNICA 2002" September 10 - 12, 2002

Professional Organizations:

Association for Research in Otolaryngology, Acoustical Society of America, International Neural Network Society