12-16-2014 12:00 AM

Effect of Reverberation Context on Spatial Hearing Performance of Normally Hearing Listeners

Renita E. Sudirga
The University of Western Ontario

Supervisor
Ewan Macpherson
The University of Western Ontario Joint Supervisor
Margaret Cheesman
The University of Western Ontario

Graduate Program in Health and Rehabilitation Sciences
A thesis submitted in partial fulfillment of the requirements for the degree in Doctor of Philosophy

Recommended Citation

Sudirga, Renita E., "Effect of Reverberation Context on Spatial Hearing Performance of Normally Hearing Listeners" (2014). *Electronic Thesis and Dissertation Repository*. 2621.
https://ir.lib.uwo.ca/etd/2621

EFFECT OF REVERBERATION CONTEXT ON SPATIAL HEARING PERFORMANCE OF NORMALLY HEARING LISTENERS

(Thesis format: Monograph)

by

RENITA ELIZABETH SUDIRGA

Graduate Program in Health and Rehabilitation Sciences
(Hearing Science Field)

A thesis submitted in partial fulfillment
of the requirements for the the degree of
Doctor of Philosophy

The School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario, Canada

Abstract

Previous studies provide evidence that listening experience in a particular reverberant environment improves speech intelligibility and localization performance in that environment. Such studies, however, are few, and there is little knowledge of the underlying mechanisms. The experiments presented in this thesis explored the effect of reverberation context, in particular, the similarity in interaural coherence within a context, on listeners' performance in sound localization, speech perception in a spatially separated noise, spatial release from speech-on-speech masking, and target location identification in a multi-talker configuration.

All experiments were conducted in simulated reverberant environments created with a loudspeaker array in an anechoic chamber. The reflections comprising the reverberation in each environment had the same temporal and relative amplitude patterns, but varied in their lateral spread, which affected the interaural coherence of reverberated stimuli. The effect of reverberation context was examined by comparing performance in two reverberation contexts, mixed and fixed. In the mixed context, the reverberation environment applied to each stimulus varied trial-by-trial, whereas in the fixed context, the reverberation environment was held constant within a block of trials.

In Experiment I (absolute judgement of sound location), variability in azimuth judgments was lower in the fixed than in the mixed context, suggesting that sound localization depended not only on the cues presented in isolated trials. In Experiment II, the intelligibility of speech in a spatially separated noise was found to be similar in both reverberation contexts. That result contrasts with other studies, and suggests that the fixed context did not assist listeners in compensating for degraded interaural coherence. In Experiment III, speech intelligibility in multi-talker configurations was found to be better in the fixed context, but only when the talkers were separated. That is, the fixed context improved spatial release from masking. However, in the presence of speech maskers, consistent reverberation did not improve the localizability of the target talker in a three-alternative location-identification task. Those results suggest that in multi-talker situations, consistent coherence may not improve target localizability, but rather that consistent context may facilitate the buildup of spatial selective attention.

Keywords: Reverberation, listening context, spatial hearing, speech intelligibility, horizontal sound localization, cocktail-party listening, spatial release from masking, perceptual adaptation

Acknowledgments

Firstly, I would like to express sincere gratitude to Dr. Ewan Macpherson and Dr. Margaret Cheesman for their guidance, support, encouragement, and patience. Their critical thoughts and constructive feedback have been immensely helpful throughout all stages of this project. It has been a great learning experience to work with both of you. I am also thankful for the insights, feedback, and support from Dr. Prudence Allen.

I would also like to extend my appreciation to professors, staff, and colleagues in the National Centre for Audiology and the Health and Rehabilitation Sciences program, who have created enriching and supportive environment for learning and conducting research. Without their contribution, my experience here would have been much less rewarding.

I am grateful for the opportunity to discuss this work with members of my examination committee: Drs. Vijay Parsa, Susan Scollie, Ingrid Johnsrude, and Pavel Zahorik. Thank you for your time, careful reading, thoughtful questions and feedback.

I also would like to thank my lab colleagues, Rufina Taylor, Ioan Curca, Janet Kim, Tran Nguyen, Iman Kazdagli, Parvaneh Kabirrah, for being always available for help, discussion, or chats, and also to friends who have volunteered much of their time. Special thanks also go to the study participants.

To Lenko, thank you for the editing and proofreading help, but most importantly, thank you for your love and support. I am also grateful to my family for their support and understanding during this process.

I also would like to acknowledge the financial support of AUTO21, Ontario Research Fund, Western University, and Natural Sciences and Engineering Research Council.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation

Reverberation, the collection of acoustic reflections in a room, can have positive or negative influences on how a sound source is perceived, depending on the task being performed. The addition of reverberation energy can increase the detectability of a sound source. The energy of the early part of reverberation can be beneficial for improving speech intelligibility (Bradley, Sato, & Picard, 2003; Arweiler & Buchholz, 2011). For sound localization, reverberation degrades the interaural cues used to perceive the direction of a sound source. In general, listeners perform more poorly in determining the horizontal and vertical directions of a sound source in rooms than in the free field (Hartmann, 1983; Ihlefeld & Shinn-Cunningham, 2011). For speech perception, reverberation reduces the depth of the modulation present in the amplitude envelope of natural speech, a factor that is closely related to speech intelligibility (Steeneken & Houtgast, 1980), and introduces self masking (Nábělek, Letowski, & Tucker, 1989; Nábělek & Robinette, 1978).

Results from categorical speech experiments (Watkins, 1991, 2005a, 2005b; Watkins

& Makin, 2007; Watkins, Raimond, & Makin, 2011; Aspeslagh, Clark, Akeroyd, & Brimijoin, 2014) and speech perception experiments (Brandewie & Zahorik, 2010; Srinivasan & Zahorik, 2013, 2014) indicate that listeners can adapt to the negative effects of reverberation when the reverberation in the stimulus context (e.g., stimuli in recent trials or the speech immediately preceding the target speech) matches the reverberation of the test words. In a speech categorization task (Watkins, 2005a, 2005b), listeners were asked to identify test-word tokens along a "sir-stir" continuum as either a "sir" or a "stir". The test words were constructed by modifying the amplitude envelope of a "sir" waveform, applying attenuation of various depths to introduce a gap in the waveform. An increase in the modulation depth caused the modified waveform to be perceived towards a "stir". Applying reverberation increased the proportion of test words identified as "sir" because reverberant energy masked the silent gap in the test word's stop consonant. However, the results also showed that the proportion of reverberant test words identified as "stir" increased when the reverberation in the preceding sentence carrier matched that of the test word, whereas more "sir" responses were reported when the carrier's reverberation did not match that of the test word (Watkins, 2005b) or when the reverberant test words were presented in isolation (Watkins & Raimond, 2013). In light of these results, the researchers suggested that listeners might be informed by the reverberation tails in the context, and that they might use this information to compensate for the masking effect of the reverberation on the silent gap in the test word's stop consonant (Watkins, 2005b). Information about the reverberation tails might be gained from speech offsets or from the offsets present in auditory filters during spectral transitions.

Similar evidence of perceptual adaptation to reverberation has also been shown in a speech perception task (Srinivasan & Zahorik, 2014) utilizing PRESTO (Gilbert, Tamati, & Pisoni, 2013), an open-set sentence-based speech corpus. Srinivasan and Zahorik (2014) tested word recognition performance in consistent and inconsistent reverberation contexts. In the consistent context, the same reverberation was applied to all sentences within an experimental block, whereas in the inconsistent context, the reverberation applied to each sentence was randomly chosen from a set of six different simulated rooms. Furthermore, the effects of amplitude envelope and fine-structure cues on perceptual adaptation were separately examined. To examine the effect of amplitude envelope, listeners were presented with speech stimuli with anechoic fine-structure cues modulated by the reverberant speech envelope. To examine the effect of fine-structure cues, the fine structure in the reverberant speech was modulated with the anechoic speech envelope. Better speech intelligibility scores were obtained in consistent reverberation context, but only for the stimuli containing reverberant amplitude envelopes. This result was consistent with the findings in the categorical speech experiments (Watkins, 2005a, 2005b), which indicated that the perceptual adaptation could be attributed to the matched reverberation in the stimulus' amplitude envelope.

Several studies have also demonstrated similar perceptual adaptation for speech-in-noise tasks (Brandewie & Zahorik, 2010; Zahorik & Brandewie, 2011; Srinivasan & Zahorik, 2013). In these studies, the target sentence was presented from the front while a stationary Gaussian noise was presented from the right of the listener. The results from these studies showed that listeners obtained better speech intelligibility scores when the sentences were presented in a consistent reverberation context. Using closed-set speech materials from the Coordinate Response Measure corpus (Bolia, Nelson, Ericson, & Simpson, 2000), Brandewie and Zahorik (2010) found a 2.7-dB lower speech reception threshold, which amounted to an approximately 18%-improvement in word-recognition score, when the reverberation context was consistent.

The perceptual adaptation observed in the categorical speech experiment (Watkins, 2005b) was found to be more robust under monaural listening. However, this is in contrast to the adaptation observed in the speech-in-noise experiment (Brandewie & Zahorik, 2010). In the latter study, a statistically significant effect of context (i.e., adaptation) was observed under binaural listening, but not when listeners performed the task monaurally. It was suggested that the mechanism responsible for the effect observed in the speech-in-noise study was different from that contributing to the adaptation in the categorical speech experiment (Brandewie & Zahorik, 2010; Watkins & Raimond, 2013). Watkins (2005b) demonstrated that the adaptation observed in their experiment could be mainly attributed to the amplitude modulation in the context. On the other hand, Brandewie and Zahorik (2010) suggested the contribution of the *buildup of the precedence effect* to explain the perceptual adaptation observed in their speech-in-noise experiment, based on the time scale of the perceptual adaptation (within a few seconds following an initial exposure) and its binaural nature.

Because reverberation takes time to build up, the onset of a reverberant sound is typically dominated by the direct sound, and therefore contains reliable spatial cues. The buildup of the reverberant energy causes more degradation in the later part of the sound. The precedence effect (Wallach, Newman, & Rosenzweig, 1949; Zurek, 1980; Blauert, 1997; Litovsky, Colburn, Yost, & Guzman, 1999; Shinn-Cunningham, Zurek,

& Durlach, 1993, see also Sec. 2.3.2 for an overview) refers to a binaural process that emphasizes the spatial information at the onset. Furthermore, when listeners are repeatedly presented with a reverberant stimulus, the perceptual weight of the onset spatial cues is increased relative to the spatial cues in the later part of the signal. The latter phenomenon is commonly referred to as the buildup of the precedence effect (Freyman, Clifton, & Litovsky, 1991; Yang & Grantham, 1997; Djelani & Blauert, 2001; Blauert & Col, 1992).

Studies of the buildup of the precedence effect have mostly employed simple stimuli, such as clicks or noise bursts, or have been conducted under the simplified case of a single reflection. Only a few studies have utilized complex stimuli, e.g., speech and music (Djelani & Blauert, 2001) or multiple reflections (Djelani & Blauert, 2001; Zahorik, Brandewie, & Sivonen, 2009). A study investigating the precedence effect in a diffuse stationary noise found a substantial weakening of the precedence effect (Chiang & Freyman, 1998). However, the buildup of the precedence effect in the presence of other competing sound sources has not been yet investigated. Besides the study of Brandewie and Zahorik (2010), there is little research on the relationship between the buildup of the precedence effect and our abilities to perform spatial hearing tasks in daily listening situations, which often involve complex and dynamic stimuli, reverberation, and competing sound sources.

Additionally, the majority of research on the effect of reverberation on auditory spatial perception has mainly examined the relationship between reverberant acoustic cues and the observed behavioral responses. There exists little research on how spatial perception is affected by *variability* in the reverberation context. The aim of the current research is to examine the effect of reverberation context on spatial hearing abilities of normally hearing listeners in a variety of listening tasks common in everyday listening situations.

1.2 Objectives

The studies presented in this thesis shared a common goal of evaluating the effect of consistency in reverberation context on spatial hearing abilities of normally hearing listeners. Additionally, we aimed to examine whether the perceptual adaptation phenomenon described in the previous section can be attributed to compensation for degraded interaural coherence in the reverberant stimuli.

Participants were tested in their abilities to perform spatial hearing tasks under mixed and fixed exposure conditions. Simulated rooms (reverberation) were used in all of the experiments. There were three simulated rooms, denoted as R1, R2, and R3. Additionally, the anechoic free field was denoted as R0. In mixed exposure conditions, the simulated room was randomly chosen on each trial, whereas in fixed exposure conditions, the simulated room was consistent across all trials within an experimental block.

The reverberation characteristics of the simulated rooms were particularly designed to examine the role of interaural coherence in perceptual adaptation to reverberation. All simulated reverberation had a similar temporal envelope but the spread of the lateral reflections was varied. An increase in the spread of lateral reflections causes interaural coherence to degrade. The potential effect of the reverberation's amplitude envelope on perceptual adaptation (Watkins, 2005a, 2005b; Srinivasan & Zahorik, 2014) was controlled by keeping the amplitude envelope of the reverberation similar from room to room. The method used to generate the simulated reverberation and characterizations of the simulated reverberation are presented in Chapter 3.

The specific objectives of the studies presented in this thesis are:

1. To test the hypothesis that listeners can localize a reverberant sound source more accurately when they receive consistent reverberation exposure.

2. To examine the effects of interaural coherence in the reverberant stimuli and the consistency in the reverberation context on speech intelligibility performance in the presence of a spatially separated noise.

3. To test the hypothesis that spatial release from speech-on-speech masking in reverberation is better in a consistent than an inconsistent reverberation context.

4. To test the hypothesis that the ability to localize a speech target in the presence of similar competing speech is better in a consistent than an inconsistent reverberation context.

5. To examine whether improved localizability of the target talker may contribute to the effect of reverberation context in the speech-on-speech task (speech perception in a multi-talker listening situation).

1.3 Significance

Further understanding of mechanisms that contribute to perceptual adaptation in reverberation can be useful for several applications. Knowledge of aspects of reverberation (e.g., reverberation time, interaural coherence), as well as their limits, that allow perceptual adaptation to occur may be used to improve processing strategies in assistive hearing devices. Preservation or maintenance of stimulus features that positively contribute to perceptual adaptation may help listeners to overcome the detrimental effects of reverberation in challenging listening situations. Furthermore, as discussed previously, studies of auditory spatial perception in reverberation have rarely taken into account context effects. Thus, information about the amount of perceptual adaptation that can occur in various listening situations may potentially be used to improve the accuracy of methods that predict listeners' performance in reverberation.

The studies presented in this thesis will specifically evaluate the effect of degradation in interaural coherence on perceptual adaptation. To our knowledge, no previous studies have examined this aspect of reverberation in relation to perceptual adaptation.

1.4 Document overview

Chapter 2 includes a literature review of relevant areas and introduces the concepts and notations that will be used in the subsequent chapters. The acoustic characterization of reverberation is formally discussed. A review of acoustic cues for sound localization and the perceptual weighting of those cues is presented, followed by a discussion of the effects of reverberation on the acoustic cues and the localization performance of normal-hearing listeners. A perceptual mechanism that facilitates robust sound localization performance in reverberation (i.e., the precedence effect) is described. Furthermore, the role of spatial hearing in speech perception in the presence of energetic and informational maskers and the effects of reverberation on spatial release from masking are discussed. The chapter concludes with a summary and with a discussion of the relationship between the current knowledge and the studies presented in this thesis, as well as further rationale behind each hypothesis.

Chapter 3 presents the experimental setup, stimuli, and procedures common to Experiments I, II, and III. The method employed to generate the simulated reverberation, the characteristics of the simulated reverberation, and the acoustic localization cues resulting from the simulated reverberation are also presented.

Chapter 4 describes a localization experiment that explored the effect of consistency in reverberation context on sound localization performance (Experiment I). Normally hearing listeners were asked to localize a two-word speech target under mixed and fixed reverberation contexts. Overall, the variance in listeners' localization responses was smaller in the fixed than in the mixed reverberation context when the stimuli were reverberant (i.e., in R1, R2, and R3). The main result of this experiment provided support for the hypothesis that stimulus history affects sound localization responses. Less variability in the reverberation context yielded better localization performance. A discussion of a possible cue re-weighting mechanism is also presented.

Chapter 5 describes a speech-in-noise experiment that explored the effects of interaural coherence and consistency in reverberation context on speech intelligibility performance (Experiment II). The experimental setup was similar to that employed by Brandewie and Zahorik (2010). The target speech was presented to the front of the listener while a steady-state noise was presented to the right of the listener, and speech perception performance was measured in consistent and inconsistent reverberation contexts. The main effect of context was significant. However, the interaction between reverberation context and room (i.e., degree of interaural coherence) was found to be not statistically significant. That is, the effect of reverberation context was similar in all rooms, including in R0 (free field). Given the similarity of the reverberation time in all the reverberation used, this result suggested that consistency in reverberation could not alleviate the degradation in interaural coherence.

Chapter 6 presents two experiments that examined speech perception (Experiment IIIA) and target location identification (Experiment IIIB) performance in multi-talker listening situations in mixed and fixed reverberation contexts. In both experiments, speech from three same-sex talkers was simultaneously presented to the listeners. In the speech perception task (Experiment IIIA), the talkers were either co-located (all talkers at 0°) or spatially separated (a target talker at 0° and the two competing talkers at -22.5° and +22.5°). Listeners were asked to identify the target words spoken by the target talker. The result indicated that speech intelligibility performance was better in the fixed reverberation context, but only when the talkers were spatially separated. This suggested that the benefit of talker separation in reverberation was facilitated by a consistency in the reverberation context.

In Experiment IIIA, listeners were informed of the talker spatial configuration (co-located or separated) and the location of the target talker (useful when the talkers were spatially separated). Experiment IIIB was designed to examine the effect of reverberation context when listeners had no prior information of the target location. Three same-sex talkers were positioned at -22.5°, 0°, and +22.5°, and the listeners were asked to identify the position of the talker uttering a designated target word (the callsign “Baron” in the CRM corpus). The results showed that listeners performed similarly in mixed and fixed reverberation contexts. This suggests that the effect of reverberation context obtained in Experiment IIIA was likely not due to improved localizability of the target talker. Rather, the consistency in reverberation might facilitate listeners' ability to selectively attend to the target speech by its known location.

Chapter 7 concludes the thesis by presenting a summary of the findings obtained in all studies, of the relationships between them and to those of previous studies, and of the limitations of the studies. Further discussions on the contribution of the current research and future research directions are also presented.

Chapter 2

Background

2.1 Chapter overview

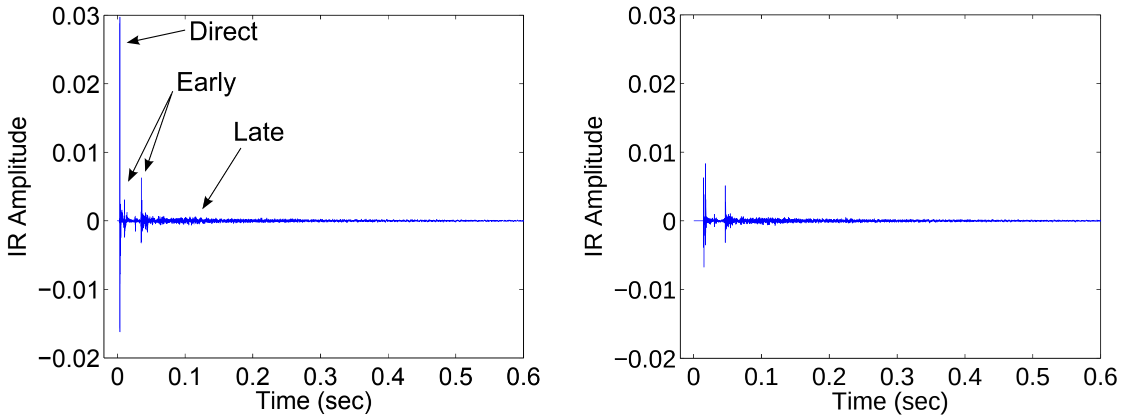
The ability to perceive the directions of sounds in space is undoubtedly essential for our spatial awareness. The acoustic cues that we use for localizing sounds, however, are often degraded by acoustical reflections from physical objects present in our surroundings (e.g., walls). In this chapter, the basic theory of sound localization and spatial hearing processes involved in the free field and in reverberation will be presented. It aims to provide the readers with information about the auditory processes involved in Experiments I, II, and III, as well as to introduce the notations that will often be used in the following chapters.

Section 2.2 first presents the formal characterization of reverberation and its parameters; the concepts introduced in this section will be referred to when discussing the effects of reverberation on the acoustic cues for speech and spatial perception. Section 2.3 presents the theory of sound localization in the free field, including the auditory cues for sound localization and their perceptual weighting, as well as the effects of reverberation on the auditory localization cues and localization performance.

Furthermore, the precedence effect, a perceptual mechanism which alleviate the ambiguity in auditory spatial cues introduced by reverberation, will be discussed. It has been proposed that the buildup of the precedence effect may contribute to the perceptual adaptation observed in the speech perception study of Brandewie and Zahorik (2010). Section 2.4 discusses the role of spatial hearing in alleviating the negative effects of maskers on speech perception (i.e., spatial release from masking). The concepts of energetic and informational masking are discussed, followed by the a discussion on the mechanisms involved in the spatial hearing of energetic and informational masking. Knowledge of these mechanisms will be useful for identifying the perceptual mechanisms involved in the speech-on-speech masking experiment (Experiment III). Section 2.5 discusses the effect of reverberation on spatial release from masking. Finally, Section 2.6 provides a summary of the literature reviewed, and presents further motivation and rationale for each experiment presented in this thesis.

2.2 Reverberation

When a sound propagates from its source to a receiver, its characteristics are modified by the environment in which the event takes place. Utilizing the linear system theory, the effect of the environment on the source can be characterized by its impulse response (IR), or equivalently, by its unique frequency-domain representation, the transfer function (TF). We will refer to the pair of impulse responses corresponding to the transmission paths from a sound source to a listener's two ears in the free field as the head-related impulse responses (HRIRs), and the head-related transfer functions (HRTFs) as their frequency-domain counterparts. In rooms, the corresponding



(a) Distance to source = 1 m **(b)** Distance to source = 4 m

Figure 2.1: Room impulse responses measured at the left ear of a dummy head for a frontal source at (a) 1m and (b) 4 m.

IR and TF pairs will be referred to as the binaural room impulse response (BRIR) and the binaural room transfer function (BRTF), respectively.

Fig. 2.1 illustrates two examples of BRIRs. These BRIRs were taken from the publicly available Aachen Impulse Response database (Jeub, Schafer, & Vary, 2009, http://www.ind.rwth-aachen.de/en/research/tools-downloads/); the room in this example is Aula Carolina. Both IRs shown in Fig. 2.1 were taken from the left ear of the dummy head. The source was located to the front of the dummy head at a distance of 1 m (Fig. 2.1a) or 4 m (Fig. 2.1b). The characteristics of room impulse responses can be illustrated with these examples. The initial portion (approximately up to about 50 ms) of the impulse responses is comprised of distinct impulses. The first impulse corresponds to the first-arriving wavefront (the direct sound). Its amplitude tends to be higher than the later-arriving wavefronts and decreases with increasing source distance. The following impulses correspond to the early reflections. They can be associated with particular directions, and thus, their patterns depend on the source and listener positions within the room. The late reverberation can be characterized as a random, noise-like signal with a smoothly decaying envelope. Unlike the early reflections, the late reverberation is diffuse; its energy is independent of the source distance and relatively uniform at all locations in the room.

The decay time of reverberation energy is formally quantified as its reverberation time. A common measure of reverberation time is T_{60}, which roughly corresponds to the time it takes for a steady-state sound in the room to decrease by 60 dB immediately after the source is turned off (Kuttruff, 2000).

The energy of the direct sound relative to that of the remaining part of the BRIR is commonly referred to as the direct-to-reverberant energy ratio or D/R. The D/R mainly depends on the distance from the source to the receiver and on the size and absorption characteristics of the room. The further the source, the lower is the D/R. This can be illustrated by a comparison of the amplitudes of the direct sounds in Figs. 2.1a and 2.1b; it is clear that the intensity of the direct sound for the near source (Fig. 2.1a) is higher than that for the far source (Fig. 2.1b). In room acoustics, however, the D/R is a rarely used parameter for characterizing reverberation. A closely related measure is the clarity index, C_{50}, which quantifies the relative energy of the early part of the BRIR (up to its first 50 ms) to the energy in the late part of the BRIR. C_{50} originated from the perspective of speech intelligibility. It has been shown that the energy of the early part of reverberation is beneficial for improving its signal-to-noise ratio (Bradley et al., 2003).

2.3 Sound localization

Throughout this document, we will use the following convention to describe horizontal position in space relative to the centre of a listener's head. Lateral angle ($-90^o...+90^o$) refers to horizontal angular displacement relative to the median plane. Azimuth ($-180^o... + 180^o$) refers to the angular displacement relative to the front ($0°$). That is, the lateral coordinate specifies the left/right position, while the azimuthal coordinate specifies the left/right as well as front/back position. Positive and negative signs denote positions in the right and left hemispheres, respectively.

2.3.1 Sound localization in free field

To determine the lateral position of a sound source, the auditory system relies primarily on interaural difference cues (Strutt, 1907). These cues naturally arise from the physical interaction between the acoustic wave and the listener's head and body. Interaural level difference (ILD) arises due to the head shadow, effecting attenuation of the incident sound wave at the other side of the head. Interaural time difference (ITD) is the difference between the time of arrival of the incident sound wave at the ipsilateral ear relative to the contralateral ear. Congruent to the ITD (with a modulo of 2π) is the interaural phase difference (IPD), which is the difference in the instantaneous phase of the incident sound wave at the ipsilateral ear relative to the contralateral ear.

Our ability to utilize ITD and ILD as localization cues is limited by the sensitivity of the binaural system to these cues. Due to the dimension of the head (average radius of 8.75 cm), as well as the ambiguity of IPD, the maximum frequency which permits unambiguous ITD estimation for a source located along the interaural axis is approximately 1500 Hz (Stevens & Newman, 1936; Zurek, 1980). Coincidentally, this value is close to the maximum frequency at which the just-noticeable-difference (JND) in IPD is measurable via dichotic tones presented over earphones. The JND-IPD function increases with frequency and reaches a vertical asymptote at approximately 1300 Hz (Zwislocki & Feldman, 1956). On the other hand, the ILD due to head shadow is small for sounds with wavelengths greater than the radius of the head (i.e., for low frequency sounds). ILD increases with frequency, and at approximately f >4000 Hz, it becomes a reliable cue for sound localization. Experimental data on localization error of tonal stimuli in free field (Stevens & Newman, 1936) illustrate the complementary role of ITD and ILD over different ranges of frequency. Localization error is small for low-frequency tones below 1300 Hz due to the availability of reliable ITD. Between 1300 Hz and 4000 Hz, localization error is large as neither the ITD nor ILD are reliable. Beyond 4000 Hz, localization performance improves again as the ILD becomes increasingly reliable.

Sounds that we encounter in our daily lives are rarely pure tones. Wider spectral bandwidth allows the binaural system to integrate the spatial cues across frequency to determine the apparent sound location (Stern, Zeiberg, & Trahiotis, 1988). Given a broadband stimulus containing a combination of both the ITD and ILD, the ITD carries a greater weight in determining the apparent sound location (Wightman & Kistler, 1992; Macpherson & Middlebrooks, 2002). Studies concerning the relative weighting of different spatial cues have typically been conducted employing dichotic stimulation or virtual auditory environments. Contrary to the studies performed in free field which permit only ecologically plausible combination of cues, such techniques allow for independent manipulation of different cues. In one such study (Macpherson

& Middlebrooks, 2002), an ITD or ILD bias was imposed on virtual free-field stimuli, and the listeners' localization responses to the biased stimuli were measured. When the stimulus contained low-frequency components (including broadband noise containing both low- and high-frequency components), imposing an ITD bias was more effective in shifting the localization responses than imposing an ILD bias. For high-pass filtered stimuli ($f > 4000$ Hz), the relative importance of ILD and ITD was switched. That is, an imposed ILD bias shifted the localization responses more effectively than an imposed ITD bias.

Besides the low-frequency ITD cue, which is derived from the fine structure of the binaural waveforms, the high-frequency components in the stimulus can also provide listeners with a timing difference cue arising from their interaural group delay. This high-frequency ITD cue is commonly referred to as the envelope-ITD cue (McFadden & Pasanen, 1976). Envelope ITD can be extracted at the onset of a stimulus, provided it is sufficiently abrupt, or when the stimulus contains ongoing modulation in its amplitude envelope. While listeners' sensitivity to the fine-structure ITD is relatively independent of the spectral bandwidth, the sensitivity to envelope ITD increases with increasing spectral bandwidth (McFadden & Pasanen, 1976), as an increase in bandwidth can be associated with more pronounced modulation in the signal's amplitude envelope. However, the ability to use envelope-ITD cues for sound localization varies considerably across listeners, and the perceptual weight of such cue is relatively weaker than the perceptual weights of fine-structure ITD and of ILD cues (Macpherson & Middlebrooks, 2002; Trahiotis & Bernstein, 1986). The study of Macpherson and Middlebrooks (2002) showed that listeners who benefited from envelope ITD mainly derived the information from the onset of the high-pass filtered stimulus, but the effectiveness of the envelope ITD was increased when an ongoing amplitude modulation was applied.

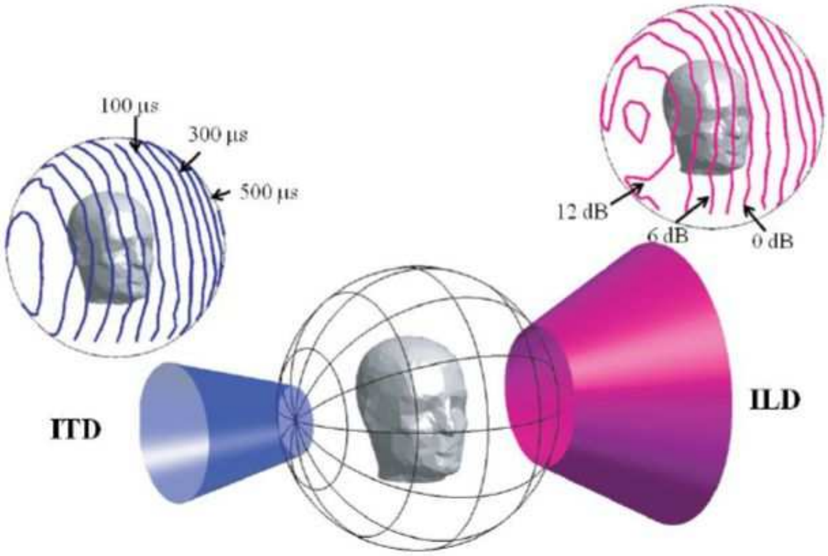


Figure 2.2: Illustration of the cone of confusion. The blue(left) and magenta (right) iso-countour lines connect all points in the sphere that have the same ITD or ILD, respectively. The corresponding cones (bottom-centre figure) contain roughly all the points in space corresponding to a particular ITD or ILD. Illustration provided by Ewan Macpherson, and originally appeared in the thesis of Kassandra-Anne Birtch (2012).

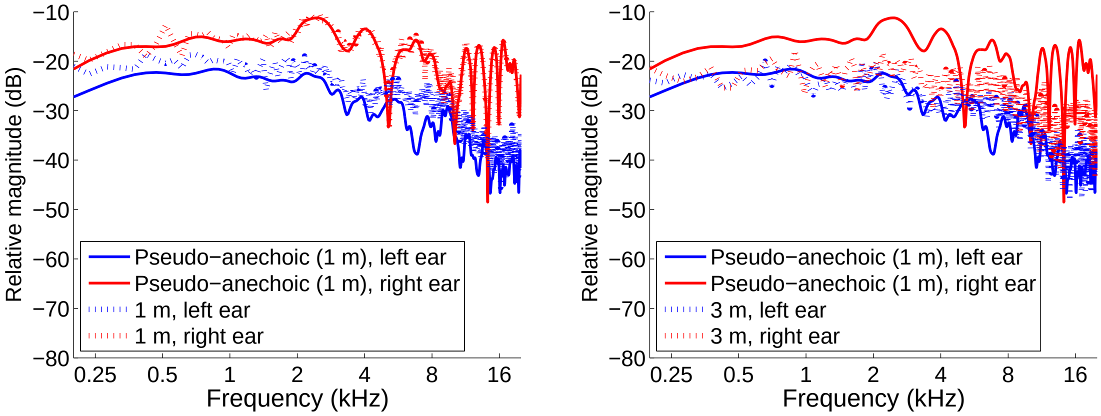
While interaural signal differences are useful to determine the lateral angle of a sound source, they do not inform us about its vertical displacement or front/back position. A particular value of interaural signal difference corresponds to a set of locations along its corresponding "cone of confusion", illustrated by the blue and magenta areas in shown in the bottom-centre figure of Fig. 2.2. Thus, additional information is needed to resolve the polarity/vertical position within a cone of confusion. Vertical and front/back orientations are primarily cued by spectral shape at the high-frequency range in the ipsilateral ear (i.e., monaural cue) (Macpherson &

Sabin, 2007), particularly the spectral peaks and notches above 5 kHz (Wightman & Kistler, 1992). The variation in HRTF spectral shape can be attributed to the pinna's transfer function, which is dependent on the direction of sound incidence (Blauert, 1997, pp.63-69). Due to the individual variation in pinna shapes, error in the vertical and front/back dimensions is particularly common when a localization experiment is conducted with non-individualized HRTFs, while the accuracy of the lateral response is generally maintained (e.g., Wanrooij & Opstal, 2005; Mendonça, Campos, Dias, & Santos, 2013).

We have omitted the discussion of another spatial dimension, namely distance. In free field, distance perception is largely dependent on the attenuation of the intensity of the source with distance, as well as the attenuation of high-frequency components due to the resistance characteristics of the air. Thus, it is largely affected by listeners' knowledge about the source. Furthermore, in contrast to perception of sound direction, room reverberation has a positive effect on distance perception as D/R also provides a cue to source distance. We will not discuss distance perception in further details as the D/R of our stimuli will be kept constant.

2.3.2 Sound localization in reverberation

The primary acoustic cues for sound localization in free field have been discussed in the previous section. Sound localization in reverberant fields is also based on the same set of cues; however, reverberation may cause conflicting or ambiguous cues. In this section, we will first describe the degradation of the acoustic cues used for sound localization, then discuss the perceptual processes that deal with the ambiguities. The flexibility of these processes allows us to perform robust sound localization in



(a) Distance to source = 1 m   **(b)** Distance to source = 3 m

Figure 2.3: Magnitude spectra of BRTFs measured at the ears of a dummy head with the source to the right of the listener (+90° azimuth). The solid lines correspond to a pseudo-anechoic BRIR at 1 m (see text), and the dashed lines to a (a) source at 1 m and (b) source at 3 m. BRIR data are taken from the Aachen Impulse Response database.

diverse environments.

Generally, the degradation introduced by reverberation of the cues for sound localization is comprised of decorrelation of the binaural waveforms and an increase in bias and variability in the spectral shape and the interaural signal differences. Fig. 2.3 shows a comparison between the magnitude spectra of BRTFs measured with the source at 1 m and 3 m away from the listener (Figs. 2.3a and 2.3b, respectively). These BRIR data were taken from the Aachen Impulse Response database (see Sec. 2.2, first paragraph). A pseudo-anechoic HRTF was constructed from the initial 5-ms part of the 1-m BRIR, setting the remaining part of the BRIR to zero. The measurements were taken in a stairway with the source located at +90° azimuth.

The increased frequency-to-frequency variability due to reverberant energy can be readily seen by comparing the magnitude spectra of the pseudo-anechoic and unmodified BRTFs in Fig. 2.3a. Furthermore, reverberation energy masks the notches and peaks in the spectral profile of the anechoic HRTF. ILD, indicated by the level difference between the left-ear (blue) and right-ear (red) spectrum magnitude curves in Fig. 2.3b, is significantly reduced with increasing source distance.

The effects of reverberation on the steady-state BRTF magnitude spectra and interaural signal difference is summarized in the following list (Shinn-Cunningham, Kopčo, & Martin, 2005; Rakerd & Hartmann, 1985, 2010):

1. An increase of random variation (both in time and frequency) in the BRTF magnitude spectra with increasing source distance (*cf.* decreasing D/R). Furthermore, the variability is larger at the contralateral ear.

2. A decrease in ILD with increasing source distance (the amount of ILD is limited by the D/R at the ipsilateral ear). Also, the variability in ILD increases with increasing source laterality and distance.

3. Interference between the direct sound and reflections may cause interaural cues that are not plausible in free field, for example, large low-frequency ILDs or implausibly large ITDs.

4. Reverberation decorrelates the binaural waveforms. Hence, interaural coherence, a measure of similarity between normalized binaural waveforms, is reduced with decreasing D/R.

5. The presence of asymmetric, intense early reflections may introduce a systematic bias in ITD and ILD. An example of this situation is when a listener is positioned with one ear close to a wall.

A decrease in interaural coherence implies a decrease in the validity of ITD information; the binaural system relies on the similarity of the binaural waveforms to extract interaural temporal difference. A decrease in listeners' sensitivity to ITD with decreasing interaural coherence has been demonstrated by Rakerd and Hartmann (2010); sensitivity to the ITD of a 750-Hz narrowband noise in a room producing an interaural coherence of 0.2 is close to chance performance (as measured with a left-right discrimination task). Such low value of coherence can be found in real rooms. Listeners in the study of Hartmann (1983) exhibited poor performance in localizing a 500-Hz tone in rooms without an attack onset, i.e., in the absence of reliable ITD cues. However, their performance improved with increasing spectrum bandwidth and density. That is, an increasingly reliable location estimate can be obtained when spatial cues across frequency are combined.

Rakerd and Hartmann (1985) also demonstrated that while listeners could not localize a 500-Hz tone in the absence of an attack transient in a room, their response followed a systematic pattern, and for certain source locations, followed the low-frequency ILD pattern. For a far source, large low-frequency ILDs are implausible in free field, although for a nearby source, such cues indicate its distance. Note that in free field, in which salient ITD cues are available, low-frequency ILD has little effect on listeners' localization responses (Macpherson & Middlebrooks, 2002). Results from a time-intensity trading study which pitted the effectiveness of ITD against ILD (Rakerd & Hartmann, 2010) demonstrated a decrease in ITD/ILD trading ratio (i.e., an increase in the effectiveness of ILD in shifting lateralization responses) with a decrease in interaural coherence. These results illustrate that in complex environments, in which spatial cues are differentially degraded, various spatial cues may be weighted according to their reliability. The dominance of low-frequency ITD, as that found in free field, holds to a certain extent in reverberation. That is, low-frequency

ITD is weighted more heavily than other interaural cues. However, when its reliability becomes increasingly degraded, other interaural cues (ILD or envelope ITD) may become increasingly effective in shifting localization responses (Nguyen, 2014).

A broadband noise without an abrupt onset can be successfully localized in a room, but the accuracy of the responses is dependent upon reverberation time (Hartmann, 1983; Giguère & Abel, 1993). An increase in reverberation time implies a decrease in D/R as well as increased variability and bias in the BRTF spectral shapes and in the interaural difference cues. The average RMS-errors in the location identification of a 500-Hz tone with a slowly-gated onset in (Hartmann, 1983) are 2.3° and 3.2° in the absorbing and reflecting rooms, $T_{60} = 1$ s and $T_{60} = 5.5$ s at 1000 Hz respectively. With stimuli containing attack transients, however, reverberation time has little effect on localization accuracy. With 50-ms, 500-Hz tonal pulses as stimuli, listeners in the same study attained on average RMS-errors of 3.3° and 3.4° in the absorbing and reflecting rooms, respectively. In the following section, the importance of onset cue in sound localization is discussed.

2.3.3 The precedence effect

Because reverberation energy takes time to build up, the presence of a salient, abrupt stimulus onset can provide listeners with reliable spatial cues within a brief duration prior to the buildup of the reverberation energy. Both psychophysical (Wallach et al., 1949; Zurek, 1980) and physiological evidence (Devore, Ihlefeld, Hancock, Shinn-Cunningham, & Delgutte, 2009; Litovsky & Yin, 1998; Litovsky et al., 1999) indicates that the perceptual and physiological systems weight onset spatial cues more heavily than the spatial cues at the later part of the stimulus, particularly when the ongoing spatial cues are ambiguous. This perceptual phenomenon is well known as the precedence effect (Wallach et al., 1949; Zurek, 1980). Historically, the term "precedence effect" has also been used to refer to several phenomena believed to stem from the same mechanism (Litovsky et al., 1999).

The precedence effect can be viewed as a means to resolve the ambiguity that may present in the ongoing spatial cues (Zurek, 1980). Its strength can be demonstrated in a well known illusion called the Franssen effect (Hartmann & Rakerd, 1989). In this illusion, two loudspeakers placed to the left and right of the listener emit identical tones. The left loudspeaker first plays the tone with an abrupt onset. The left tone is rapidly gated off with a decaying level as the right loudspeaker plays the tone with a rising level. Listeners localize the tone to the left loudspeaker even though it has been the right speaker that plays the tone most of the time. Furthermore, most listeners do not notice any transition from the left to right loudspeakers. It was later found that the room was a necessary component for the Franssen illusion to work. Hartmann and Rakerd (1989) demonstrated that the Franssen illusion failed in an anechoic environment in which all steady-state cues were salient or when the stimuli had dense spectra. Similarly, in lateralization experiments, Freyman, Zurek, Balakrishnan, and Chiang (1997) demonstrated that the dominance of the onset cues only held when the ongoing cues were ambiguous. When reliable information could be obtained from the ongoing spatial cues, the lateralization judgement was dominated by the latter cues, even when salient cues were available at the onset.

The dominance of onset spatial cues in determining the apparent sound location is commonly referred to as the *localization dominance* aspect of the precedence effect.

Also associated with the precedence effect are the *echo suppression* or *fusion* and *lag-discrimination suppression* phenomena (Litovsky et al., 1999; Shinn-Cunningham et al., 1993; Litovsky & Shinn-Cunningham, 2001). Demonstrations of these phenomena typically utilize a simplified case of echoic stimuli consisting of pair of spatially-separated leading and lagging tokens. The lead and lag tokens are separated in time with an inter-stimulus interval (ISI). If the ISI is sufficiently small, listeners perceive a single sound event (echo suppression); its apparent location depends on the ISI. For an ISI smaller than the temporal range of binaural summation (e.g., between 0-1 ms for broadband clicks), the apparent location can be determined from the interaural difference cues resulting from the superposition of the lead and lag waves at the ears. Beyond this range (e.g., between 1-10 ms for broadband clicks), the lead position dominates the apparent location (localization dominance). As the ISI is increased to reach the *echo threshold*, two distinct events are perceived. They can be perceived as originating either from the same position (towards the direction of the lead) or from different locations. The apparent direction of the lag token tends to be perceived towards the direction of the lead token (lag-discrimination suppression).

A neurophysiological correlate of onset dominance has been shown in the low-frequency, ITD-sensitive neurons in the inferior colliculus (IC). In these neurons, the discharge rate increases monotonically with azimuth, particularly to the side contralateral to the measurement site. Devore et al. (2009) measured the response of a sample population of neurons to reverberant 200-ms broadband noise. Data from the study indicated better directional sensitivity during the earlier part of the reverberant stimulus (rate averaged across the 50-ms duration re: stimulus onset) in the neuron population measured. The measurement taken during the later part of the stimulus indicated a compressed rate vs. azimuth function (i.e., worse directional sensitivity) compared to that taken in the earlier part of the stimulus. In general, the population firing rate is proportional to the interaural coherence; coherence is high at the onset as the direct sound dominates the ear-input signals and low for the later part of the stimulus. A neurophysiological correlate of echo suppression has also been observed in the IC. Using the lead-lag stimulus pair of clicks or tone bursts, Litovsky and Yin (1998) measured the responses of neurons in the same site (IC) to the lagging click or noise (also see Litovsky et al., 1999). The data suggested a suppressed response to the lag stimulus (re: response to the same stimulus presented in isolation) that recovered with an increasing ISI. The recovery time (delay corresponding to a lag-stimulus response rate at least half for that stimulus presented in isolation) ranged from 1.5 to 154 ms with a median of 27 ms for the population of neurons studied.

The time scale of echo suppression is considerably larger than the recovery time from monaural forward masking, which is in the range of 2-4 ms for click stimuli as measured in the ventral cochlear nucleus and the auditory nerve (as cited by Litovsky et al., 1999). Furthermore, echo suppression does not suppress the intensity of the lagging sound. Rather, its effect is limited to suppressing its directionality (Freyman, McCall, & Clifton, 1998). Besides facilitating sound localization in reverberation, the fusion aspect of the precedence effect may also indicate its role in the formation of auditory objects in reverberation.

Two aspects of the precedence effect, namely echo suppression and lag-discrimination suppression, have been shown to increase in strength with stimulus repetition (Freyman et al., 1991; Yang & Grantham, 1997; Djelani & Blauert, 2001). Using click stimuli, the echo threshold for a single lead-lag click pair is about 5 ms, whereas presenting a train of 9 lead-lag click pairs at a rate of 4 pairs/s increases the echo threshold to 26 ms (Yang & Grantham, 1997). The lag-discrimination threshold for the same stimuli is 5 ms for the single click pair and about 10 ms for the train of click pairs. Interestingly, single-neuron responses in the IC do not exhibit buildup under repeated stimulus presentation (Litovsky & Yin, 1998).

The buildup of echo suppression has been interpreted as an indication of recalibration of the auditory scene as listeners accumulate more reliable evidence from the repeated onsets (Clifton, Freyman, & Meo, 2002). Left-right asymmetry has also been observed for echo threshold, with a greater echo threshold obtained when the lead stimulus originates from the right side (24.4 ms and 18.6 ms for right- and left-lead stimuli, respectively, measured with short noise bursts, Grantham, 1996). Litovsky et al. (1999) noted that the behavioral measures of the precedence effect suggest processes that are also mediated in the auditory pathway higher than the level of the brainstem, suggesting the contribution of the auditory cortex. Furthermore, both exogenous (stimulus-driven) and endogenous (goal-oriented) spatial attention have been shown to significantly affect the echo threshold (Spence & Driver, 1994; London, Bishop, & Miller, 2012).

The effect of repeated stimulus presentation on localization dominance, however, has been much less explored. Using lead-lag stimuli (123-μs rectangular pulses) presented under headphones, Brown and Stecker (2013) measured listeners' lateralization responses under single- or repeated-stimulus conditions, using an experimental paradigm similar to that employed by Yang and Grantham (1997). Lateralization was induced by manipulating either the ITD or ILD. While a buildup of fusion (increase in echo threshold) was observed under the repeated-stimulus condition, the influence of the lag position on apparent lateralization also increased with increasing ISI, and no sign of increased dominance of the lead position was observed. The authors suggested that the increase of suppression associated with the buildup in fusion serves primarily to strengthen the grouping of the lead and lag into a single auditory object rather than to enhance localization. Further studies utilizing more natural stimuli or different types and degrees of stimulus ambiguity must be performed to examine the role of repeated onsets on the buildup of localization dominance.

A long-term effect of stimulus exposure has also been shown to improve sound localization performance in rooms in horizontal, vertical, and distance dimensions (Shinn-Cunningham, 2000). It is unclear whether such improvement is related to the perceptual processes contributing to the buildup of onset dominance or echo suppression.

2.4 Spatial release from masking in speech perception

When multiple sound sources are present, spatial separation among sources can alleviate the detrimental effects of the competing sources on the target. This benefit of spatial separation is commonly referred to as *spatial release from masking.* The mechanisms that contribute to spatial release from masking are different depending on how the maskers affect the target, i.e., whether the masking is energetic or informational in nature. The following subsections first introduce the concepts of energetic and informational masking, followed by the mechanisms involved in spatial release from energetic and informational masking.

2.4.1 Energetic and informational masking

In everyday situations, speech communication is often conducted with other sound sources simultaneously present in the background. Irrelevant sound sources, which we will refer to as maskers, can be harmful to the perception of the target sound in two different ways. First, the energy of the target and maskers may occupy the same spectral range; thus exciting the same neural population along the auditory pathways, starting from the auditory periphery. Second, the presence of maskers can increase the complexity of the auditory scene, rendering it more difficult to perceptually organize the various components in the sound mixture into different auditory objects (object formation and segregation). Such an increase in complexity also increases the demand on our cognitive resources, making it more difficult to direct and maintain our attention to the target sound. The terms *energetic masking* and *informational masking* are commonly used to refer to masking components that predominantly affect the perception of target by the first and second means, respectively (e.g., Freyman, Helfer, McCall, & Clifton, 1999).

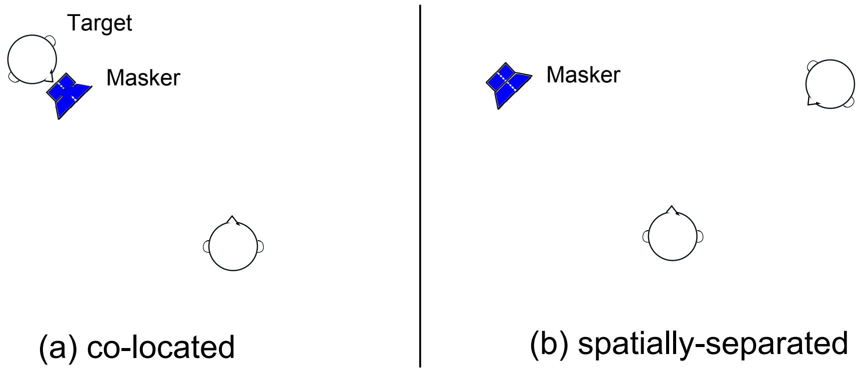
A masker likely affects the target perception in both ways, and it may not always be feasible to isolate the effects of energetic and informational components of masking. Functionally, the amount of energetic masking is quantified by utilizing maskers that possess little similarity to the target, thus making it relatively easy to identify the target. For example, the perception of familiar speech in a stationary noise can be considered as being mainly affected by energetic masking; the two signals have very different characteristics, and furthermore, we are highly familiar with the acoustic characteristics of speech. In this case, the perception of the target (e.g., its intelligibility) is mainly affected by the detectability of the target energy across different frequency bands. Articulation-index based models of speech intelligibility (French & Steinberg, 1947; ANSI, 1997) have been able to predict behavioral speech intelligibility performance rather successfully when the maskers are mainly energetic in nature (Kryter, 1962). Such prediction is essentially calculated from the signal-to-noise ratios (SNRs) across independent critical bands and the speech band-importance function, a function which weights the contribution of each particular band to the overall speech intelligibility.

Informational masking is closely related to difficulty in the processes involved with auditory scene analysis (Bregman, 1990). Contrary to energetic masking, informational masking may occur even when there is no significant overlap between the target and masker representations in the auditory periphery. Two factors have been identified as contributing to informational masking: 1) target and/or masker uncertainty, and 2) target-masker similarity (Lutfi, Chang, Stamas, & Gilbertson, 2012; Ihlefeld & Shinn-Cunningham, 2008); higher stimulus uncertainty and similarity of the target and masker attributes (e.g., pitch, timbre, or location) contribute to increased informational masking. To a certain extent, the Gestalt principle governs the grouping of sound components into auditory objects; sound components are automatically grouped according to the similarity and continuity of their attributes in time and frequency. Multiple objects with various saliency in their attributes compete to be in the auditory foreground. Exogenous factors, such as contextual information, familiarity with the target sounds, and selective attention, also take part in the perceptual process, enhancing the representation of the object in the foreground and further enhancing the segregation process (Kahneman, 1973).

An example of common listening situations involving informational masking is the cocktail-party setting in which target speech is masked by speech from other talkers. In this situation, the listener must selectively attend to the speech spoken by a target talker. Besides the relative energy of the target to the masker, performance in this task is highly dependent on the similarity between the target and competing talker's voice. In an experiment conducted by Brungart, Simpson, Ericson, and Scott (2001), speech intelligibility performance in a diotic listening condition was measured with the CRM stimuli in the case of one competing talker when the masker was either same-sex, different-sex, or the same talker. Performance was best when the competing talker was of a different sex, and worst when the competing talker was the same as the target talker. Additionally, a modulated speech-shaped noise with the same amplitude envelope as the speech stimuli was used to compare the effect of energetic masking. In the latter case, performance was better than when speech is used as a masker. Similar results were also obtained by Festen and Plomp (1990); SRT in the case of reversed speech was dependent on the talker sex. Talker sex, however, had a much smaller effect on SRT when the masker was a speech-shaped noise with a spectrum matched to that of the competing speech.

2.4.2 Spatial release from energetic masking

Spatial release from energetic masking is attributed to two known factors: the head shadow effect and binaural unmasking (Freyman et al., 1999). The advantage due to the head shadow effect, commonly referred to as the *better-ear advantage*, is illustrated in Fig. 2.4. In the co-located condition, the effects of the head shadow on the target and the masker are equal. In the spatially separated condition, the target and the masker signals are differentially affected by the head. Thus, if there is spatial separation, one ear receives a favorable signal-to-noise ratio, except when the target and masker are placed symmetrically about the interaural axis. Comparing the spatial configuration in the co-located and spatially separated conditions, illustrated in Fig. 2.4a and 2.4b, respectively, in the spatially separated case the right ear receives a better signal-to-noise ratio (SNR) due to the the head attenuating the masker's energy. The better-ear advantage attributed to head shadow, relative to the frontal co-located condition, amounts to approximately 5-10 dB for a masker located at one side of the interaural axis (+/-90°) (Bronkhorst & Plomp, 1988)); the SNR required to obtain a given level of performance was 5-10 dB lower when the masker was located at +/-90° rather than at 0°.



(a) co-located

(b) spatially-separated

Figure 2.4: Illustration of the better-ear advantage. In (a) the co-located condition, both ears receive the same SNR. In (b) the spatially-separated condition, the right ear receives a better SNR due to the head shadow effect.

The advantage due to binaural unmasking, commonly referred to as *binaural squelch*, is mainly attributed to the target and maskers' differing in their fine-structure ITDs (Bronkhorst & Plomp, 1988; Hirsh, 1950); thus, it mainly operates on the low-frequency components of the signals. The increase in effective SNR can be successfully modeled with equalization-cancellation processing on the binaural signals (Durlach, 1963). While the better-ear advantage arises mainly from the physical effect of the head shadow on sound propagation (and thus, it is essentially dominated by the SNR at the ear that receives the better SNR), binaural unmasking requires binaural input. Thus, binaural squelch is often referred to as the true binaural advantage. For normally-hearing listeners, binaural squelch attributed to ITD alone amounts to 4-5 dB for sentence intelligibility in the presence of a stationary noise located at 30° to 150° relative to the frontal co-located configuration (Bronkhorst & Plomp, 1988).

Articulation-index based speech intelligibility models that account for better-ear advantage and binaural squelch (e.g., Jelfs, Culling, & Lavandier, 2011; Lavandier & Culling, 2010) can predict the effect of target and masker spatial separation relatively well when the masker mainly contains energetic masking components. However, they tend to overestimate performance when the masker is also speech, and are lacking in mechanisms capable of predicting the effects of talker similarity on listeners' performance.

2.4.3 Spatial release from informational masking

Spatial release from informational masking occurs when a perceived difference in spatial attributes provides the listener with additional information about the identity of the target and/or the masker. When there exist other robust cues that can also differentiate the target speech, such as differences in pitch, vocal-tract length, or intonation, adding a spatial separation may not result in a significant improvement in intelligibility (Darwin & Hukin, 2000a). If other attributes are similar, however, spatial separation can provide a robust cue to segregate the speech mixture. The benefit of spatial separation in a speech intelligibility experiment with the CRM stimuli with two competing, same-sex talkers positioned symmetrically at ±90° around the

0° target talker amounts to approximately 12 dB relative to the frontal co-located condition (Marrone, Mason, & Kidd., 2008).

An example that illustrates the different roles of spatial separation in alleviating the effects of energetic and informational masking in speech perception is presented in the study by Freyman et al. (1999). The experiment utilized stimuli consisting of lead-lag stimulus pairs separated by 4 ms. The lead and lag stimuli were presented from two loudspeakers located at 0° and +60°. In the "FR-FR" condition, the lead stimuli of the target and the masker were presented from the front loudspeaker. 4-ms following the lead, the lag stimuli were presented from the right (+60°) loudspeaker. In the "FR-RF" condition, the target's lead stimulus was presented from the front loudspeaker and its lag was presented from the right, while the masker's lead stimulus was presented from the right and its lag was presented from the front. Switching the lead and lag loudspeakers did not appreciably affect the long-term spectra of the lead-lag stimuli. However, due to the precedence effect, the FR-stimulus was perceived to be at the front while the RF-stimulus was perceived to be at the right. Using this experimental configuration, the intelligibility of female speech was measured in the presence of a masker. The masker was either a speech-shaped noise (energetic masking) or another speech sample spoken by a different female talker (informational masking). Consistent with the prediction that energetic masking can be mainly explained by the amount of overlap in the target's and masker's energy, the speech intelligibility in the FR-FR condition did not differ significantly from that in the FR-RF condition when the speech-spectrum noise acted as a masker. For the speech-on-speech masking, however, the intelligibility in the FR-RF condition was significantly better than that in the FR-FR condition.

Hawley, Litovsky, and Culling (2004) examined the effects of increasing the number of interferers on spatial release from energetic and informational masking. Speech-spectrum noise and speech from the same talker as the target were used to represent energetic and informational masking, respectively. Adding a number of interferers, particularly when they were distributed more evenly around the listener, increased the masker energy at the better ear as well as decorrelating the binaural input, reducing both the better-ear advantage and binaural squelch. Spatial release in speech-on-speech masking, however, was robust in all spatial configurations tested (all maskers in the same position or spatially distributed across the hemifields); the proximity of the target to the interfering sources played a more important role than the number of interferers.

Spatial release from informational masking can be partly attributed to the role of spatial cues in selective attention. Attention can either be driven by salient attributes of the selected object (stimulus-driven) or be volitionally directed based on the task demand (task-driven). To examine the role of spatial attention in resolving speech-on-speech masking, Kidd, Arbogast, Mason, and Gallun (2005) manipulated listeners' expectation of the target location in a multi-talker speech perception task. Three CRM sentences spoken by different same-sex talkers were simultaneously presented to the listeners at equal level. The talkers were located at -60°, 0°, and +60°, and the target sentence was presented from one of these locations with various degrees of certainty. In the condition with 100% certainty, $p = 1$, the target was always presented at the same location (-60°, 0°, or +60°), while in the condition with the least certainty, $p = 0.33$, the target location in each trial was chosen randomly with equal probability. When $p = 1$, target recognition was close to maximum (92%) for all target locations, indicating that each sentence could be fully segregated from the mixture. Consistent with the ability to direct attention, a decline in performance was observed with decreasing certainty in target location, with performance reaching 67% on average in the least certain condition.

Selective attention is comprised of two different aspects, namely reorientation and maintenance (Kahneman, 1973, pp. 119). Maintenance of attention is associated with the ability to continuously attend to the selected message; for example, in a task in which the listener is required to attend and verbally repeat a message played to one ear while ignoring the message played to the other ear, e.g., (Cherry, 1953). Reorientation of attention is associated with a switch in the identity of the target message; for example, when a target speech is played with random switches from one ear to the other. It is known that task performance is impaired following an attentional switch but that it improves over time (Kahneman, 1973, pp. 119). That is, it takes some time for selective attention to build up. Furthermore, continuity in the attended attribute over time facilitates refinement in the spatial tuning of selective attention.

The build up of spatial selective attention and the cost of switching attention can be illustrated in the study of Best, Ozmeral, Kopčo, and Shinn-Cunningham (2008). In this study, listeners were required to identify a sequence of target digits; each digit was presented simultaneously with 4 other spatially-separated competing digits. The location of the target digits were fixed or randomized. A set of LED lights informed the listeners of the location of each of the target digits, either synchronously with or in advance (up to 1 s) of the presentation of the digits. The results showed that listeners performed better in the fixed target location condition (i.e., more correct digits were recalled). Providing an advanced visual cue of the target location reduced, but did not eliminate the cost of switching location.

The tuning of spatial release from informational masking is explored by Marrone et al. (2008). In this experiment, speech intelligibility in a multi-talker situation was tested with the CRM stimuli. Three CRM sentences spoken by different same-sex talkers were presented simultaneously to the listener. The target talker was always positioned at 0°, and the two maskers were symmetrically positioned to the left and right of the target talker at ±15°, ±45°, ±90°(spatially separated condition), or with both maskers co-located with the target at 0°. Spatial release from masking was measured as the difference between the speech reception thresholds in the co-located and spatially separated conditions. Results from the study indicated that the benefit of spatial separation increased with the amount of spatial separation. The benefit was particularly apparent when the spatial separation was small, but the increase in benefit became smaller with increasing separation. The amounts of spatial release from masking were 8.7 dB, 12.3 dB, and 12.6 dB for 15°, 45°, and 90° spatial separation, respectively. That is, the largest increase in masking release was obtained when the spatial separation was increased from 0° to 15°, and no further significant increase was obtained by increasing the spatial separation from 45° to 90°.

2.5 Effect of reverberation on spatial release from masking

Spatial release from masking can be significantly reduced in reverberation. As discussed in Sec. 2.3.2, reverberation decorrelates the interaural signals. Consequently, this reduces the effectiveness of the binaural unmasking process. Better-ear advantages can also be significantly reduced, as the reflections increase the energy of the masker at the ear contralateral to the target source. The reduction in spatial release from energetic masking attributed to reverberation is illustrated in the study of Kidd, Mason, Brughera, and Hartmann (2005); speech perception in co-located (0°,0°) and spatially separated (0°,+90°) configurations was measured with 6-8 band vocoded CRM speech materials presented simultaneously with noise containing frequency bands overlapping those of the speech (same-band noise masker condition in the study). The target was always presented at 0°. In the least reverberant condition, spatial release from masking amounted to about 8 dB, while it was reduced to 2 dB in the most reverberant condition. Marrone et al. (2008) showed that the amount of spatial release from energetic masking (approximated using reversed speech materials) was reduced from 4.1 dB in the absorbent room to 2.3 dB in the reflecting room for ±90° spatial separation. Recall that the study utilized three sound sources with the maskers symmetrically placed on either side of the target source (0° target and ±90° maskers). The D/Rs and reverberation times were 6.3 dB and 0.06 s in the absorbent room, and -0.9 dB and 0.25 s in the reflecting room, respectively.

Reverberation also reduces the effectiveness of auditory cues that can be used to perceptually segregate competing speech (Culling, Hodder, & Toh, 2003; Lee & Shinn-Cunningham, 2008). Culling et al. (2003) showed that the effectiveness of differences in $F0$ and talker location in facilitating the recognition of target speech in a two-talker mixture was significantly impaired in reverberation. Reverberation can reduce the benefit of talkers' spatial separation in the following ways. First, the increase in the total energy of the competing speech reduces the portion of the mixture that is dominated by the target's energy (containing accurate spatial information about the target). Second, it is known that spatial difference is a relatively weak cue for perceptual grouping of simultaneous sound components (Darwin & Hukin, 1999). It is, however, an effective sequential segregation cue. It is necessary for different sources to be heard as separate streams for an accurate extraction of their distinct binaural information (Best, Gallun, Carlile, & Shinn-Cunningham, 2007). Thus, the ability to access binaural information is affected by the strength of other perceptual segregation cues. Negative effects of reverberation on other segregation cues can significantly impair listeners' abilities to extract reliable spatial information of individual sources.

While the precedence effect can be advantageous in maintaining perceived separation, steady-state cues also affect apparent sound locations. Results from Culling et al. (2003) and Darwin and Hukin (2000b) demonstrate a reduced effectiveness of spatial cues on selective spatial attention in a speech-on-speech masking task. The same studies also demonstrate reduced effectiveness of other segregation cues, namely, pitch, prosody, and vocal-tract length cues. Furthermore, reverberation may also impair listeners' abilities to perceive source locations by increasing the total masker energy; if the maskers contain amplitude modulation, reverberation fills up these gaps. During the temporary periods when the masker energy dips, and if the reverberation is low, a portion of the signal mixture is dominated by the target sound, thus allowing listeners to better estimate its identity, including its location, e.g. as the model proposed by Faller and Merimaa (2004).

2.6 Summary of literature review and its relation to the current research

The literature review presented in this chapter discussed the acoustic cues and auditory processes that are involved in the spatial hearing tasks performed in Experiment I: single-source sound localization (Sec. 2.3), Experiment II: speech perception in a spatially separated noise (Sec. 2.4.2), and Experiment III: speech perception and localization in the presence of multiple talkers, i.e., informational masking (Sec. 2.4.3).

There is evidence that when listeners have listening experience in reverberation, their performance in the particular reverberant environment can improve. This has been shown for single-source sound localization tasks in all spatial dimensions (horizontal, vertical, and distance) (Shinn-Cunningham, 2000) and for speech perception in a spatially separated noise (Brandewie & Zahorik, 2010; Zahorik & Brandewie, 2011; Brandewie & Zahorik, 2013; Srinivasan & Zahorik, 2013). These results indicate that the pattern of reverberant stimuli presented in the recent past can influence listeners' responses; however, such studies are relatively few and relatively little is known about the mechanisms underlying the changed in performance. The studies presented in this thesis aim to find further evidence of reverberation context effects in spatial hearing tasks, as well as to explore whether changes in performance could be specifically related to compensation to degradation in interaural coherence caused by reverberation.

To address the latter issue, we conducted the experiments in a simulated reverberation environment which was designed such that all reverberation maintained a similar reverberation time but varied in interaural coherence. This was performed by employing a set of loudspeakers to simulate the reflections and varying the spread of the lateral reflections while maintaining a similar temporal envelope ($T_{60} = 0.45$ s) in all reverberant conditions (Chapter 3).

Experiment I aimed to find evidence of reverberation context effects in a single-source sound localization task in the simulated reverberation environment. It is known that a repeated presentation of reverberant stimulus with salient onsets can increase the weighting of the onset cue, i.e., the buildup of the precedence effect. However, it is unknown whether a similar cue re-weighting process can also incorporate information in past trials. In Experiment I, evidence for a reverberation context effect was examined by comparing horizontal sound localization performance in mixed and fixed reverberation contexts.

Experiment II aimed to examine whether a mechanism related to compensation to degradation in interaural coherence may contribute to the enhancement in speech perception found in several studies (Brandewie & Zahorik, 2010; Zahorik & Brandewie, 2011; Brandewie & Zahorik, 2013; Srinivasan & Zahorik, 2013). Several studies have indicated the role of the amplitude envelope (see Chapter 1). To control for the potential contribution of the amplitude envelope in the reverberant context, we utilized a set of reverberation patterns which had the same reverberation time but varied in the spread of lateral reflections.

Experiment III examined the effect of reverberation context in a multi-talker listening situation (spatial release from informational masking). The effects of reverberation context on spatial selective attention and on the ability to identify the location of a target talker were examined in two experiments (IIIA and IIIB).

Chapter 3

General Methods

3.1 Chapter overview

Experiments I, II, and III were conducted using a similar experimental setup, and shared common methodology. This chapter describes these commonalities. The simulated reverberation used in all of the experiments will also be described. Methods specific to each experiment will be further described in each corresponding chapter.

The following sections describe the demographics of the study participants, experimental setup, stimuli, the method used to generate the reverberation, and analyses of the sound localization cues resulting from the simulated reverberation.

3.2 Participants

All participants were young normally hearing adults with ages between 18 and 35 years old; the majority of them were university students. Their pure-tone audiometric thresholds at 250, 500, 1000, 2000, 4000, and 8000 Hz were less than 20 dB HL (re: ANSI S3.6-1996). In total, 37 individuals participated in Experiments I, II, and III, and some individuals took part in more than one experiment. The numbers of participants in Experiment I, II, IIIA and IIIB were 8, 10, 11, and 10, respectively. An additional 9 participants performed two short experiments in Experiment IIIB. Additional information about the number of participants in each experiment and overlap of participants among experiments is presented in each corresponding chapter.

The studies were approved by Western's Research Ethics Board. All participants gave informed consent prior to their inclusion into the study and were paid $10/hour.

3.3 Equipment and setup

All experiments took place in a 5.5 m x 7 m x 3.7 m hemi-anechoic chamber. The floor was covered with sound absorbing panels and foam sheets. At the center of the room was an array of 16 loudspeakers (Tannoy i5 AW) positioned along a horizontal circle with a 1.45-m radius. The speakers were equally spaced by 22.5°. During an experiment, each participant was positioned at the centre of the loudspeaker array on a height-adjustable platform. The platform height was adjusted for each listener so that his/her ears were at the same level as the loudspeakers. Fig. 3.1 depicts this configuration.

The control computer and playback systems other than the loudspeakers were situated outside of the anechoic chamber. The stimuli were presented via an Echo AudioFire 12 audio interface (sampling rate = 44100 Hz) connected to 8 Soundweb 9008 networked signal processors, each routing an output channel of the audio interface to one of the 16 loudspeakers. The loudspeakers were driven by QSC CX168 power amplifiers. All loudspeakers were equalized to achieve flat (± 0.5 dB) frequency responses (20-kHz upper-frequency cutoff). Their gains were equalized such that a white noise sample presented from each of the loudspeakers had the same level at the

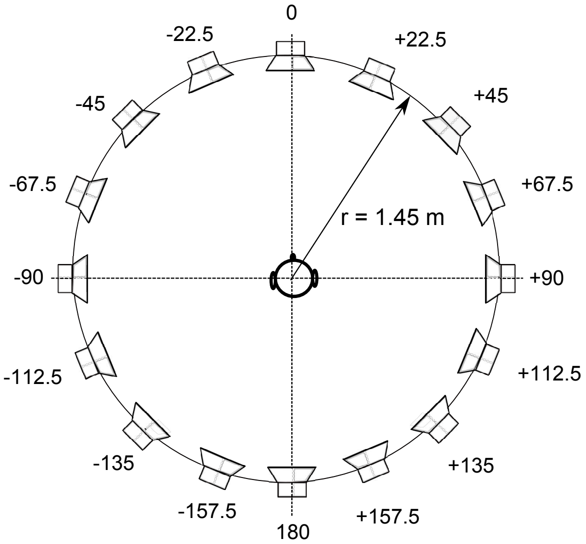


Figure 3.1: Configuration of the loudspeaker array and the listener's position in all of the experiments.

centre of the loudspeaker array. The longest inter-channel delay due to the Soundweb processors was less than 1 ms. Signal calibration was performed with a B&K 2250 sound level meter; its microphone (B&K 4189) was suspended at the center of the speaker array at the loudspeakers' height. For all stimuli, including the reverberated stimuli, the reported sound pressure level (SPL) corresponded to the measured sound level in the absence of a listener's head. Processing of stimuli and data collection were performed using custom software implemented in MATLAB (MathWorks Inc., C.A.).

3.4 Stimuli

All experiments utilized speech materials from the Coordinate Response Measure (CRM) corpus (Bolia et al., 2000). Each sentence in the corpus has the form of "Ready ⟨*callsign*⟩ go to ⟨*colour*⟩ ⟨*number*⟩ now". There are 2048 sentences in total in the corpus, corresponding to all possible combinations of eight talkers (four male and four female), eight callsigns (Arrow, Baron, Charlie, Eagle, Hopper, Laker, Ringo, Tiger), four colours (red, blue, green, white), and eight numbers (1-8). The high-frequency cutoff of the recording is approximately 8000 Hz.

A short version of the CRM sentences, used in Experiments I and II, was derived by extracting the colour and number portions of the CRM sentences with the callsign Charlie. The word boundaries were determined by visually observing their waveforms as well as by listening verification. Sound editing was performed with a custom program implemented in MATLAB. The extracted target words were gated with 1-ms raised cosine onset and offset ramps. All extracted waveforms were scaled to the same RMS-value.

The stimulus level at the listeners' position (including the level of the reverberated stimuli) was calibrated by measuring the slow-weighted SPL of all speech tokens (including all talkers and callsigns) concatenated in a randomized order with no pauses between tokens.

3.5 Simulated reverberation

To examine the role of interaural coherence in the exposure effect discussed in the Introduction chapter (Chapter 1), the simulated reverberant environments used in this study were designed to have a similar reverberation time but vary in their interaural coherence. A reverberation time of $T_{60} = 0.45s$ (broadband) was used in the present study. This was chosen to match the broadband reverberation time of one of the rooms used in a previous related study which yielded an exposure effect (Brandewie

& Zahorik, 2010). However, in contrast to the latter study, which employed frequency-dependent reverberation time, in our study, no frequency-dependent weighting was applied to the reflections (i.e., reverberation time was equal across frequency).

The reverberation was simulated by the means of the loudspeaker array. All reflections in the reverberation were delayed and attenuated copies of the dry stimulus. We constructed three simulated "rooms", denoted as R1, R2, and R3. Additionally, the anechoic field was denoted by R0. Each room was associated with a set of reverberation patterns corresponding to different source-to-listener configurations (i.e., source directions). The simulated rooms, R1, R2, R3 differed in the width of the lateral spread of the reflections. For a particular source-to-listener configuration, the interaural coherence decreased from R1 to R3. However, the temporal pattern and relative amplitudes of the reflections, measured at the location of center of the listener's head, were the same for all reverberation patterns.

A stylized depiction of the temporal pattern and relative amplitudes of the reflections for all reverberation patterns is shown in Fig. 3.2. The first line in the figure at 0 ms represents the direct sound. Each following line in the figure represents an early reflection and the triangle represents collectively the late reverberation. Let us define this function as the reflection pattern, $r_{\theta,R_i}(n)$. For each combination of source direction and room (e.g., a 22.5° source in R2), a corresponding reflection pattern $r_{\theta,R_i}(n)$ was generated and stored.

The early and late portions of the reverberation were simulated in different manners, similar to the method used in (Bradley & Soulodre, 1995). Let us define the reflection patterns corresponding to the early and late parts of the reverberation as $r^{Early}_{\theta,R_i}(n)$ and $r^{Late}_{\theta,R_i}(n)$, respectively.

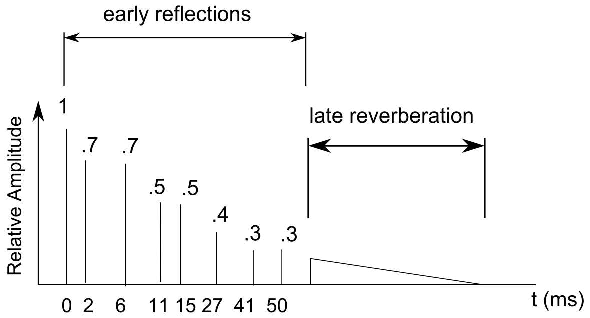


Figure 3.2: Temporal pattern and relative amplitudes of the reflections for all reverberation. Each line in the figure represents a reflection; its magnitude relative to the first-arriving sound (A_m) is shown next to it. For each reverberation, the polarities of all reflections were randomized.

For the early part of the reverberation, each reflection originated from a loudspeaker in the array. Each early reflection was an attenuated and delayed copy of the dry stimulus, and the polarities of each reflection were chosen randomly (±1) for each reverberation pattern. The active loudspeakers from which the reflections came depend on the source direction and room. Table 3.1 shows the list of active loudspeakers for each source direction and room. In R1, all reflections came from the same loudspeaker as the direct sound. In R2, the active loudspeakers (loudspeakers producing the reflections) were the loudspeakers adjacent to the direct sound as well as the direct-sound loudspeaker itself. In R3, all reflections came from -45°, +45°, -135°, and +135° regardless of the source directions. An illustration of the configuration of the active speakers for source directions of 0° and -45° in R1, R2, and R3 is shown in Fig. 3.3.

Let us define $\delta(i;\phi)$ to be the following function

$$\delta(i;\phi) = \begin{cases} 1, & \text{if } i = 0 \\ 0, & \text{otherwise} \end{cases} \tag{3.1}$$

Table 3.1: The direction of the j-th reflection ($j = 1 \cdots 7$) in the simulated reverberation for a source at θ and room R_i, $\phi_{\theta,R_i}(j)$, in degrees: (a) in R_1, (b) in R_2, (c) in R_3.

(a) $\phi_{\theta,R_1}(j)$

Direct sound direction, θ (deg)	Reflection number, j
	1 - 7
$\{-180, -135, -90, -67.5, \cdots, +67.5, +90, +135\}$	θ

(b) $\phi_{\theta,R_2}(j)$

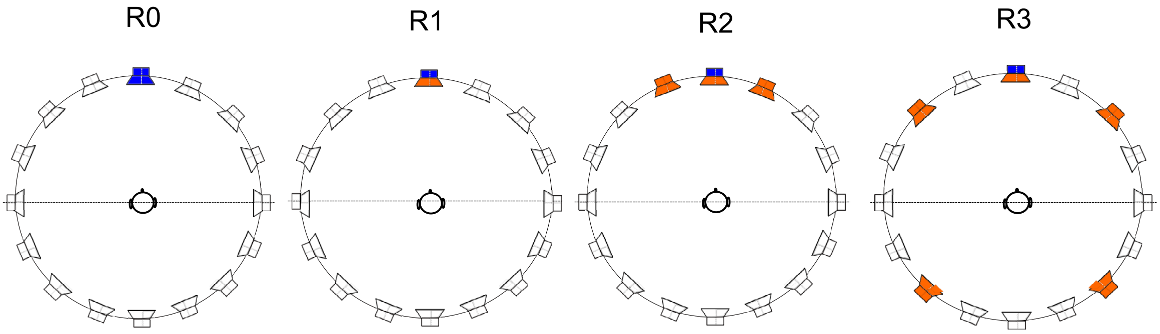
Direct sound direction, θ (deg)	Reflection number, j		
	1, 4, 6	5	2, 3, 7
0	-22.5	0	+22.5
-22.5	-45	-22.5	0
+22.5	0	+22.5	+45
-45	-67.5	-45	-22.5
+45	+22.5	+45	+67.5
-67.5	-90	-67.5	-45
+67.5	+45	+67.5	+90
$\vdots$			

(c) $\phi_{\theta,R_3}(j)$

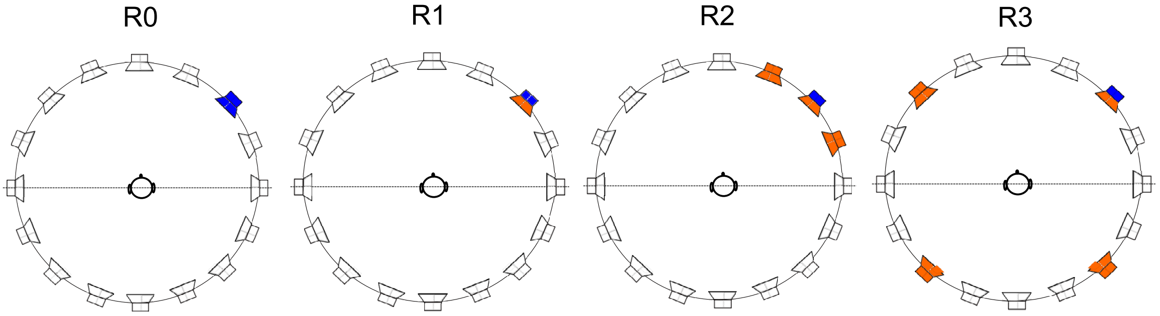
Direct sound direction, θ (deg)	Reflection number, j				
	5	1, 6	2, 7	3	4
$-180, -135, -90, -67.5, \cdots, +67.5, +90, +135$	θ	+45	-45	+135	-135

where ϕ denotes the location of the loudspeaker in degrees. Let us also define the set of active loudspeakers associated with a source at θ and room R_i as $\mathbf{\Phi}_{\theta,R_i}$. Hence, $\mathbf{\Phi}_{\theta,\mathrm{R}_1} = \{\theta\}$, $\mathbf{\Phi}_{\theta,\mathrm{R}_2} = \{\theta - 22.5, \theta, \theta + 22.5\}$, and $\mathbf{\Phi}_{\theta,\mathrm{R}_3} = \{\theta, -45, +45, -135, +135\}$.

The reflection pattern of the early reverberation corresponding to a source direction, θ, and room, R_i, can be expressed as follows.



(a) Source at 0°



(b) Source at +45°

Figure 3.3: Configuration of the active loudspeakers in R0, R1, R2, and R3 for (a) a source at 0° and (b) a source at +45°. Loudspeakers marked with blue correspond to the direct sound and those in orange correspond to the reflections.

Reflection pattern of the early reverberation:

$$r_{\theta,R_i}^{Early}(n) = \begin{cases} \sum_{m=1}^{7} (-1)^{b_m} \cdot A_m \cdot \delta(n - n_m; \phi_{\theta,R_i}(m)), & \text{if } n < 50 \times 10^{-3} \cdot f_s \\ 0, & \text{otherwise} \end{cases} \tag{3.2}$$

b_m's are independent realizations of a Bernoulli random variable with $p = 0.5$ (i.e., its outcome is 0 or 1 with equal probability). $n_m = \left\lfloor \frac{\Delta t_m}{fs} \times 10^{-3} \right\rfloor$, denotes the time-of-arrival of the m-th reflection in samples, where $\lfloor \cdot \rfloor$ denotes the operation for rounding to the lower nearest integer. A_m and Δt_m are the amplitude and time-of-arrival of the m-th reflection in ms, relative to the direct sound ($A_0 = 1$ and $\Delta t_0 = 0$), respectively.

Their values are shown in Fig. 3.2. That is, $A_{1-7} = 0.7, 0.7, 0.5, 0.5, 0.4, 0.3, 0.3$ and $\Delta t_{1-7} = 2, 6, 11, 15, 27, 41, 50$, in ms. f_s is the sampling frequency in Hz.

The late part of the reverberation, represented by the stylized triangle in the figure, was simulated with a method similar to that used by Watkins (2005b). Each impulse in the late reverberation was equally spaced by 176 samples ($\sim$ 4 ms). The polarities of the impulses were randomly chosen (± 1) for each reflection pattern corresponding to a source direction and a room. Each active loudspeaker contributed equally in magnitude and their polarities were set to be equal. The amplitude envelope of the impulses in the late reverberation was shaped with an exponential function, $\exp(-kn)$, where k is the decay rate of the exponential function. k was set to a constant such that $T_{60} = 0.45$ s (i.e., $k = 15.4$). The pattern of the impulses in the late reverberation modeled the decaying amplitude envelope and spectrum of a reverberation tail.

The reflection pattern of the late reverberation corresponding to a source direction, θ, and room, R_i, can be expressed as follows.

Reflection pattern of the late reverberation:

$$r^{Late}_{\theta,R_i}(n) = \begin{cases} 0 & ,\text{if } n < 50 \times 10^{-3} \cdot f_s \\ \frac{1}{L_i} e^{-kn} \sum_{\phi \in \mathbf{\Phi}_{\theta,R_i}} \sum_m (-1)^{b_m} \delta(n - n_m; \phi) & ,\text{if } n \geq 50 \times 10^{-3} \cdot f_s \end{cases}, \tag{3.3}$$

where L_i is the number of elements in $\mathbf{\Phi}_{\theta,R_i}$ (i.e., the number of active loudspeakers). Furthermore, the energy of the early and late $r_{\theta,R_i}(n)$ were scaled to give an early-to-late reverberation energy ratio of 14 dB (i.e., $C_{50} = 14$ dB). Let us define this scaling factor as ν.

The final reflection pattern, $r_{\theta,R_i}(n)$, for a source at θ and a room R_i was constructed by adding the direct sound, and the early and late reflection patterns (Eqns. (3.2) and (3.3), respectively),

$$r_{\theta,R_i}(n) = \delta(n;\theta) + r_{\theta,R_i}^{Early}(n) + \nu \cdot r_{\theta,R_i}^{Late}(n). \tag{3.4}$$

For each source direction and room, $r_{\theta,R_i}(n)$, was constructed and stored as a digital file. Reverberated stimuli were generated by convolving $r_{\theta,R_i}(n)$ with the dry stimuli.

In several aspects, the simulated reverberation differs from real-room reverberation. Limitations of the simulated reverberation are further discussed in Sec. 7.1.4.

3.6 Analysis of the simulated reverberation

BRIRs of the simulated reverberation were obtained with a B&K Head and Torso Simulator (HATS, type 4128). The recording system consisted of a B&K Nexus preamplifier type 2690 and the same Echo AudioFire12 audio interface used to play the excitation signal. The excitation signal used was a $2^{16} - 1$ point maximum length sequence (MLS) (Rife & Vanderkooy, 1989), approximately 1.5 s in duration. The BRTF magnitude spectra for sources at -45°, -22.5°, 0°, +22.5°, and +45° in R0, R1, R2,and R3 are shown in Fig. 3.4. The blue and red lines are the left and right BRTF spectrum magnitude data, respectively. As expected, the addition of the reflections increased the variability of the magnitude spectra and reduced the interaural level differences.

Reverberation time

The resulting reverberation times were estimated via the slopes of linear fits to the backward integration functions of the BRIR energy (Schroeder, 1965; Kuttruff, 2000),

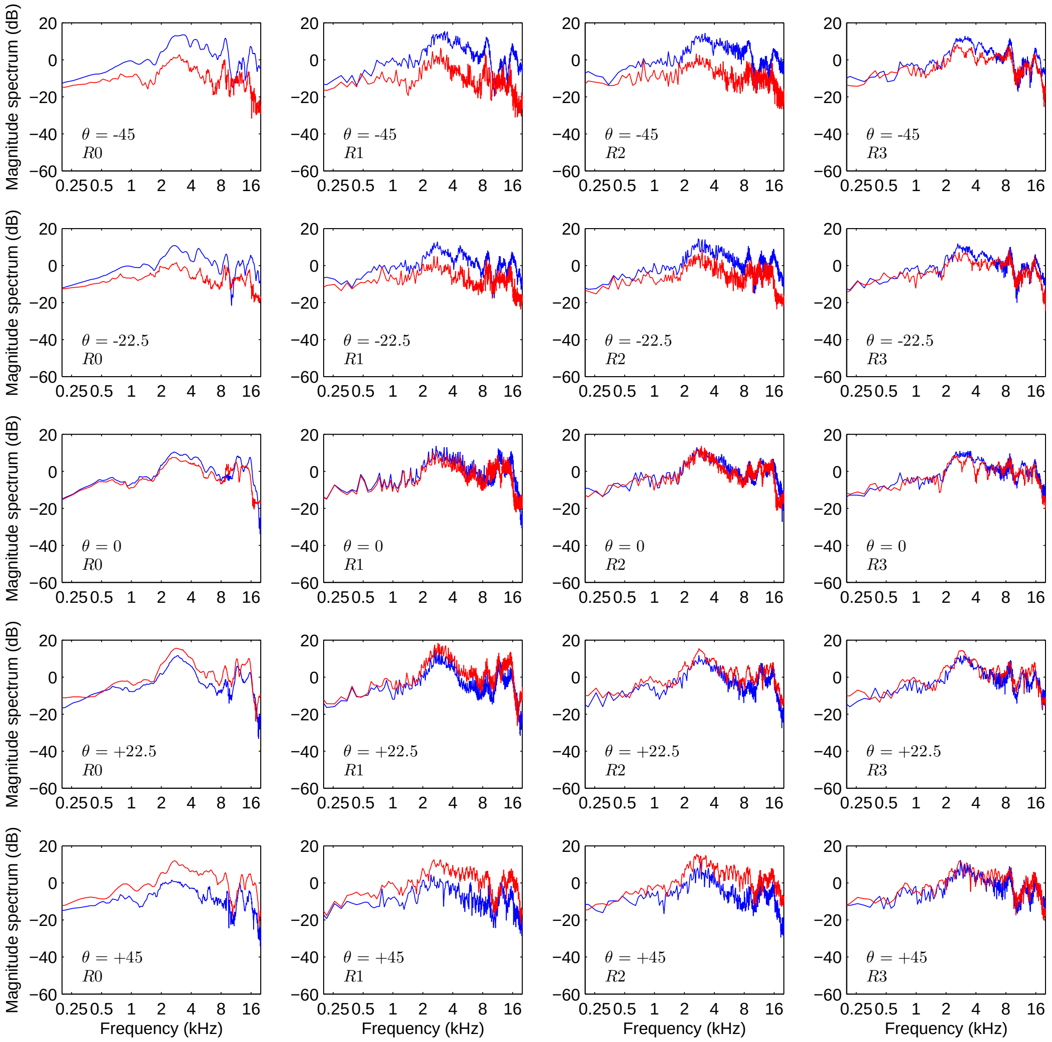


Figure 3.4: The BRTF magnitude spectra for sources at θ = -45°, -22.5°, 0°, +22.5°, and +45°in R0, R1, R2, and R3. Source direction is arranged in separate rows and room is arranged in separate columns. The blue and red lines are the left- and right-ear data, respectively.

$S(t) = \int_{t'=t}^{\infty} h^2(t')dt'$, where $h(t)$ is a BRIR. $S(t)$ is commonly referred to as the Schroeder function. The procedure to estimate the reverberation time from a BRIR is illustrated as follows. For each BRIR, the corresponding $S(t)$ was calculated; $S(t)$ corresponding to a 0° source in R3 is shown in Fig. 3.5. From each $S(t)$, a linear fit to the function was obtained from the portion of $S(t)$ that corresponded to the late reverberation (we used samples from 100-300 ms for all BRIRs); the linear fit for the left-ear BRIR in Fig. 3.5 is shown as the black dash-dot line. The reverberation time was estimated from the slope of the linear fit, $T_{60} = -60/\text{slope}$ (in seconds). The resulting reverberation times for all combinations of source positions and rooms are shown in Table 3.2. The average of all resulting T_{60} was 0.47 s with a standard deviation of 0.07.

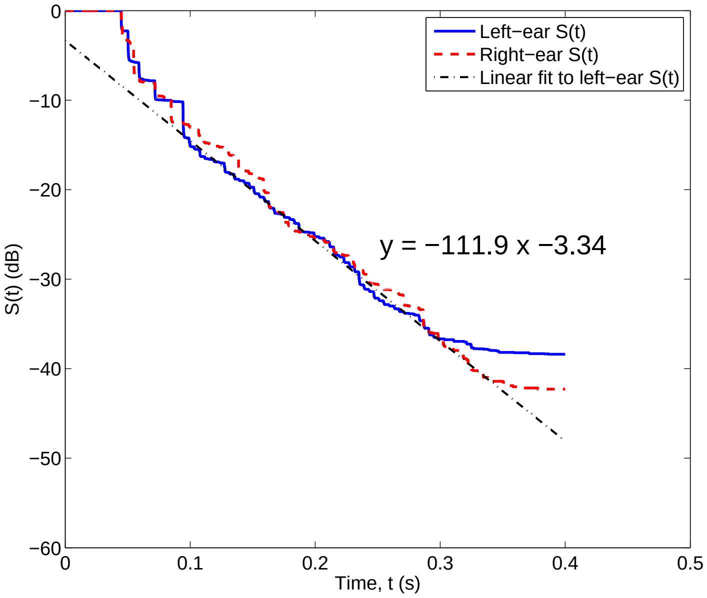


Figure 3.5: The Schroeder function, $S(t)$, of a 0° source in R3, calculated from its measured BRIR. The reverberation time was estimated from the slope of its linear fit (dash-dot line), $T_{60} = -60/\text{slope}$.

Binaural measures (ITD, ILD, IC)

Table 3.2: The measured reverberation times, T_{60}, of the simulated reverberation for all source directions in R1, R2, and R3.

Source direction (deg)	T_{60}, in seconds					
	Left ear			Right ear		
	R1	R2	R3	R1	R2	R3
0	0.56	0.46	0.55	0.55	0.50	0.52
-22.5	0.39	0.44	0.50	0.43	0.41	0.43
+22.5	0.48	0.58	0.48	0.46	0.61	0.58
-45	0.42	0.47	0.45	0.43	0.41	0.49
+45	0.39	0.56	0.43	0.44	0.56	0.43
-67.5	0.52	0.54	0.48	0.46	0.40	0.48
+67.5	0.41	0.45	0.41	0.50	0.43	0.44
-90	0.46	0.49	0.54	0.38	0.38	0.54
+90	0.42	0.45	0.47	0.57	0.46	0.54

The ILD was calculated as the ratio of the broadband energy in the right-ear BRIR to that of the left-ear BRIR.

$$ILD = 10 \log \frac{\sum_{n=1}^{N} h_r^2(n)}{\sum_{n=1}^{N} h_l^2(n)}, \tag{3.5}$$

where $h_l(n)$ and $h_r(n)$ are the broadband left- and right-ear BRIR, respectively, and N is the length of the BRIR. The ILDs are plotted in Fig. 3.6.

ITD and IC calculated from the broadband BRIR and the octave-band filtered BRIRs for centre frequencies at 500, 750, 1000, and 2000 Hz are shown in Figs. 3.7, and 3.8. ITD and IC values were estimated as follows. The interaural cross-correlation function, $IACC(\tau)$, is first calculated

$$IACC(\tau) = \frac{\sum_{n=1}^{N} h_l(n) h_r(n+\tau)}{\sqrt{\sum_{n=1}^{N} h_l^2(n) h_r^2(n)}}, \quad -10^{-3} f_s \le \tau \le 10^{-3} f_s,$$

$$-1 \le IACC(\tau) \le 1.$$

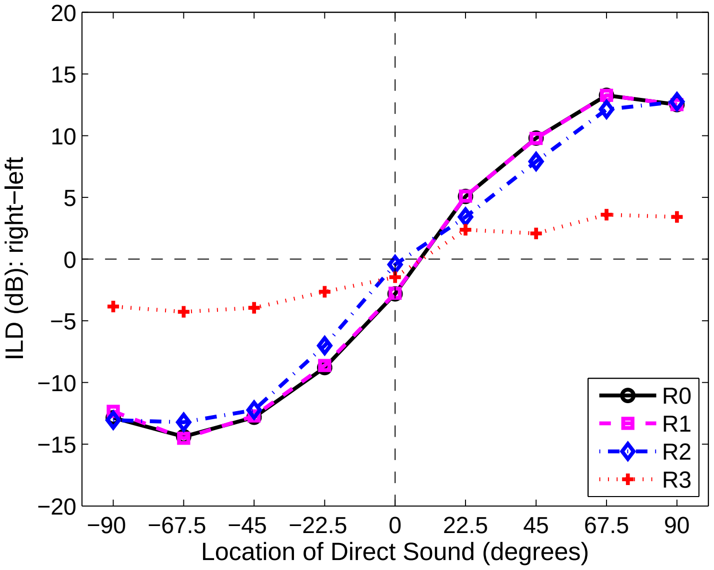


Figure 3.6: Broadband interaural level differences in R0, R1, R2, and R3.

The ITD was estimated as the time delay that corresponded to the peak of the cross-correlation function,

$$\begin{aligned} \tau^* &= \max_\tau IACC(\tau) \\ ITD &= \tau^*/f_s \quad \text{(in s)} \end{aligned} . \tag{3.6}$$

The interaural coherence, IC, was obtained as the maximum of the cross-correlation function,

$$IC = IACC(\tau^*), \tag{3.7}$$

IC was generally high in the low-frequency region and decreased with increasing frequency. In free field (R0), the IC was close to 1 and tended to decrease with increased lateral position of the source.

Inspection of Fig. 3.7 clearly indicated an increase in between-band ITD variability from R1 to R2. A significant decrease in IC from R1 to R2 was also observed (Fig. 3.8)

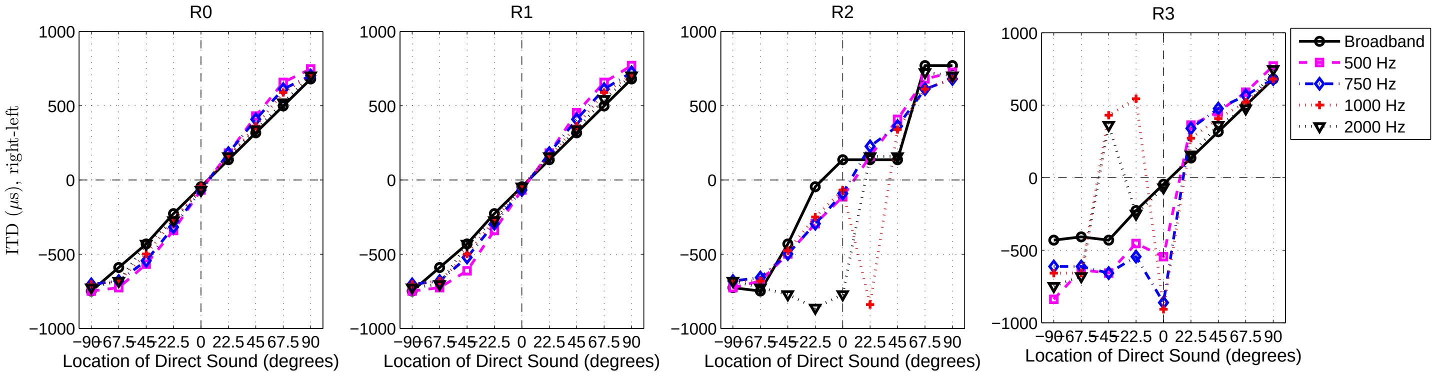


Figure 3.7: Interaural time differences in octave bands centered at 500, 750, 1000, 2000 Hz and broadband. From left to right: R0, R1, R2, and R3.

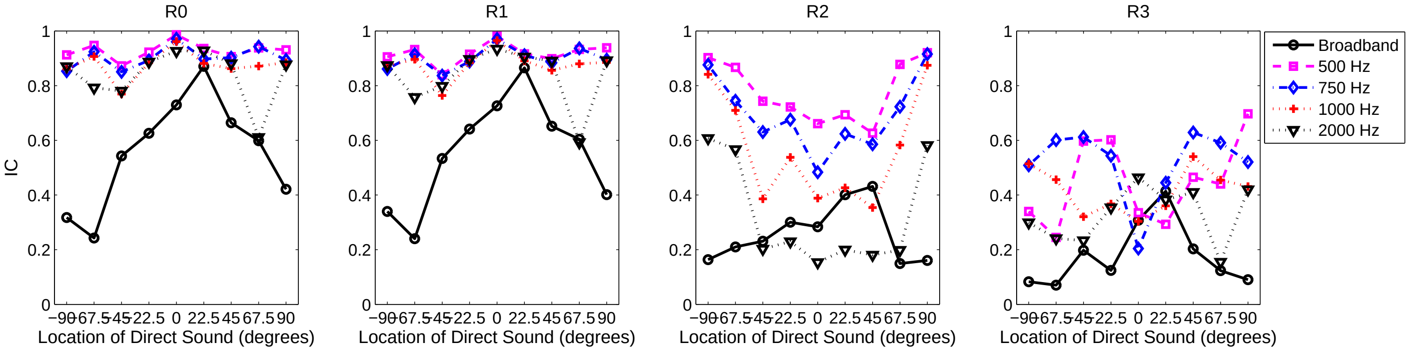


Figure 3.8: Interaural coherence in octave bands centered at 500, 750, 1000, 2000 Hz and broadband. From left to right: R0, R1, R2, and R3.

and its variability also increased. Adding reverberation with the same directionality as the direct sound (i.e., R1) did not affect either the interaural difference cues or the IC. However, an increase in variability of the magnitude spectrum can be observed (Fig. 3.4). In R3, both ILDs and ITDs were biased towards the median plane. The octave-band ICs were in the range of 0.3-0.6.

In determining the ITD of a sound with a non-zero bandwidth, the binaural system weights and combines the ITDs across frequency bands according to weighting functions that depend on frequency, IC, and ITD (Le Goff, Buchholz, & Dau, 2013).

3.7 Data analysis

The experimental data in Experiments I, II, and III were mainly analyzed with repeated-measures analysis of variance (ANOVA). The Type-I error rate, α, was set to 0.05 for all analyses. When multiple comparisons or ANOVAs were performed, the α in each test was adjusted to maintain a family-wise α of 0.05 using the Šidák correction factor. Degrees of freedom were adjusted with the Greenhouse-Geisser's ϵ; this takes into account any violation to the sphericity assumption in the data. Effect size is reported as η^2, the proportion of variance in the data that is accounted for by the model. Statistical analyses were conducted with IBM SPSS software.

Chapter 4

Experiment I: Effect of reverberation context on sound localization

In Chapter 2, the auditory cues for sound localization and the cue-weighting processes in both free and reverberant fields have been discussed. The role of stimulus context in sound localization, however, has largely been overlooked. In several other auditory tasks, for example, frequency discrimination (Demany, 1985; Amitay, Hawkey, & Moore, 2005; Watson, Kelly, & Wroton, 1976) or discrimination of interaural difference (Wright & Fitzgerald, 2001), consistency in stimulus context has been shown to have a positive effect on listeners' performance. Perhaps when localizing sounds in reverberation, listeners can accumulate information about the reverberation across trials and utilize this information to refine their listening strategies. This chapter presents a study that examines the effect of reverberation context on sound localization performance.

4.1 Introduction

In a study that examined the effect of listening experience on sound localization performance in a room (Shinn-Cunningham, 2000), it was shown that listening experience can improve the accuracy of localization judgments in azimuthal, vertical, and distance dimensions. The improvement was observed over several days of testing sessions and occurred in the absence of feedback. This indicated that the change in performance was driven by the information inherent in the stimuli. Besides this study, however, there is little research that examines the role of listening experience or the underlying mechanisms that may contribute to improved sound localization performance with experience. Evidence of listeners' taking into account stimulus pattern in the context has been demonstrated in several other auditory tasks, such as frequency discrimination of a tonal component within a tonal pattern (Watson et al., 1976), detection of tones in noise (Green, 1961; Creelman, 1973), and word discrimination under different binaural configurations (Nahum, Nelken, & Ahissar, 2010). In the last study, the ability to take advantage of the difference in IPDs of the target and the masker (referred to hereafter as the *binaural advantage*) was measured as the difference in speech reception threshold (SRT) between the $S_{\pi}N_0$ and S_0N_0 configurations. The subscript in S_{-} denotes the IPD of the masker, and the subscript in N_{-} denotes the IPD of the masker. Binaural advantage was measured under two listening conditions differing in the amount of trial-to-trial variability: in the *consistent-protocol* condition, the SRTs for the S_0N_0 and $S_{\pi}N_0$ conditions were obtained with two separate adaptive tracks; in the *mixed-protocol* condition, the SRTs were measured with two independent but interleaved tracks, the binaural configuration was randomly varied on each trial; in the *1-1-protocol* condition, the SRTs were measured with two interleaved adaptive tracks with an alternating configuration ($S_\pi N_0$-$S_0 N_0$-$S_\pi N_0$-$\cdots$). A significantly higher binaural advantage was obtained by listeners in the consistent-protocol group than that obtained by the listeners in the other groups. Furthermore, following several days of training, listeners in the 1-1-protocol group could achieve an amount of binaural advantage comparable to that achieved by the consistent-protocol group. That is, the listeners could successfully learn to discriminate $S_\pi N_0$ and $S_0 N_0$ trials within a more complex context, but it took more time for them to do so. The binaural advantage achieved by the mixed-protocol group, however, was significantly lower following the same amount of training.

Additionally, in a study that examined the effect of reverberation context in a distance perception task (Schoolmaster, Kopco, & Shinn-Cunningham, 2003), consistency of the listener's position within a room was shown to yield higher accuracy in distance judgments relative to the condition in which the listener's position was varied across trials. These studies illustrate that high variability in the context can negatively affect listeners' performance.

The present study aimed to find further evidence that the ability to localize sounds in the horizontal dimension is affected not only by the acoustic cues present in isolated trials, but also on the cues in the context. The effect of reverberation context on sound localization will be examined; listeners' ability to determine the direction of a sound source in a listening condition in which the room is consistent across trials (fixed-exposure condition) will be compared to that obtained in a condition in which trials from different rooms are presented within the same listening block (mixed-exposure condition). The simulated rooms (R0, R1, R2, and R3) presented in Chapter 3 are used; recall that the reliability of the spatial cues degraded as the lateral spread of the reflections increased (i.e., reliability decreased from R0, R1, R2, to R3). Because of a higher degree of trial-to-trial variability in the mixed-exposure condition, we expected listeners to perform better in the fixed-exposure condition.

4.2 Method

4.2.1 Participants

Eight young, normally hearing listeners participated in this study (6 females, 2 males; 20-34 years of age, mean age = 24.3). None of them had any prior experience with sound localization experiments. Total testing time per participant was 3.5-4.5 hours, divided into 3×1.5-h sessions. Testing progressed according to individual pace and schedule. Short breaks were provided whenever needed.

4.2.2 Stimuli and procedure

The experiment took place in the anechoic chamber described in Sec. 3.3. Participants stood at the centre of the circular loudspeaker array. In each trial, a short CRM stimulus (a word-and-number pair) was presented, and participants were asked to determine the direction of the sound. Stimuli were presented at 60 dB SPL with ±5-dB level rove. The lights in the anechoic chamber were turned off during the experiment, and participants performed the localization task in the dark.

Localization responses were collected via a head-tracking device (Polhemus FASTRAK) mounted on top of the listener's head. At the beginning of each trial, the listener faced the front (0°). A trial was initiated when the listener pressed the buttons on a hand-held response box. Upon detecting the button press, the program checked whether the head was correctly oriented (at 0° with ±5° and ±10° tolerance windows in the azimuthal and vertical dimensions, respectively), then presented the stimulus over the loudspeakers. The listeners oriented their heads to face the direction of the heard sound, then pressed the response-box buttons, which triggered the program to record the current head orientation position. Listeners performed the task with no feedback on the correctness of their responses.

At the beginning of the experiment, an experimenter demonstrated the procedure and trained the participants to perform the task. Listeners were given 1-2 blocks of practice trials to familiarize themselves with the task procedure. Each practice block consisted of 30 R0-stimuli that included all target directions (marked by stars in Fig. 4.1).

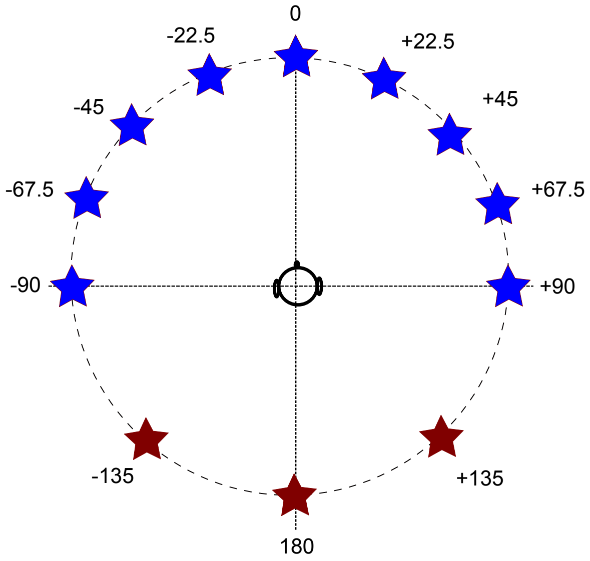


Figure 4.1: Possible target locations in the localization experiment. Numerical values are the target azimuths. Targets marked with red stars were presented so that listeners' responses were not constrained to the front hemisphere. Only responses corresponding to the front targets marked with blue stars were analyzed.

The locations of the target sounds presented to the listeners are shown in Fig. 4.1. Trials in the back hemisphere (marked with red stars) were presented so that listeners' responses were not constrained to the front hemisphere. Only responses corresponding to the front-hemisphere targets (marked with blue stars) were analyzed.

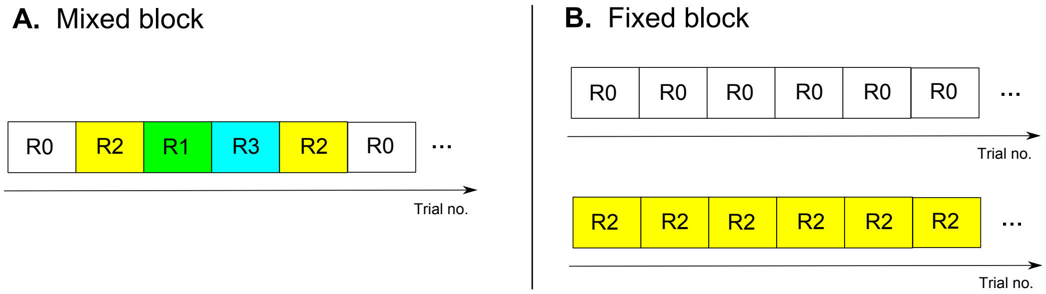


Figure 4.2: Examples of trials in (A) mixed block: trials in a block corresponded to all possible rooms (R0-R3) in a randomized order, and in (B) fixed block: all trials in this block corresponded to the same room.

Listening conditions consisted of two types of listening exposure (mixed, fixed) and four simulated rooms (R0, R1, R2, and R3). The mixed-exposure condition consisted of trials drawn from all rooms, and the order of the rooms was randomized across trials. In the fixed-exposure condition, the room was fixed across all trials within a block. Trial configurations in the mixed and fixed blocks are illustrated in Fig. 4.2. For each listening condition, each frontal-hemisphere target was presented 16 times, and each back-hemisphere target was presented 8 times. In total, there were $2 \times 4 \times (9 \times 16 + 3 \times 8) = 1344$ trials divided into 16 blocks (8 mixed-exposure blocks and 8 fixed-exposure blocks), each consisting of 84 trials.

Four participants completed the mixed-exposure condition followed by the fixed-exposure condition (Mixed/Fixed group). The remaining four participants completed the fixed-exposure condition followed by the mixed-exposure condition (Fixed/Mixed group). The order of the rooms in the fixed-exposure condition was randomized. Furthermore, both of the fixed-exposure blocks corresponding to the same room were presented consecutively within a test session.

4.3 Results

Localization response azimuths from each participant are plotted in Figs. 4.3 and 4.4. Data points plotted with red circles and blue crosses correspond to responses in mixed- and fixed-exposure conditions, respectively. Data points around the lower diagonal line correspond to responses close to the target, while those lying about the upper-diagonal line correspond to front/back reversals. It can be seen that the majority of errors were front/back reversals. Front/back reversals were rare in R0 and R1 (except for listener S24); this was expected as the pinna cue was accurate (in R0) or minimally affected by the reverberation (in R1). The number of individuals who made front/back reversal errors increased in R2 and in R3. All individuals made such errors, except S21 who only localized to the front. This is consistent with the increasing distortion and variability in the spectral shape with increasing lateral spread of the reverberation (*cf.* Fig. 3.4).

The numbers of front/back errors, defined as responses located in the back hemisphere (recall that the targets were located in the front hemisphere), are shown in the bar plots in Fig. 4.5. Front/back reversals seem to occur more often in the left hemifield than in the right hemifield. However, this pattern seems to contradict other findings which indicated more accurate vertical sound localization in the left hemifield (Giguère, Lavallée, Plourde, & Vaillancourt, 2011). Thus, our results might have been due to the specifics of our listening environments.

For the purpose of data analysis, we defined “non-reversed” responses as any frontal-hemisphere responses for targets between -45° and +45° (inclusive), and all responses for more lateral targets (±67.5° and ±90°). In this way, front-back reversals were excluded from the analysis (for lateral targets, front-back reversals are

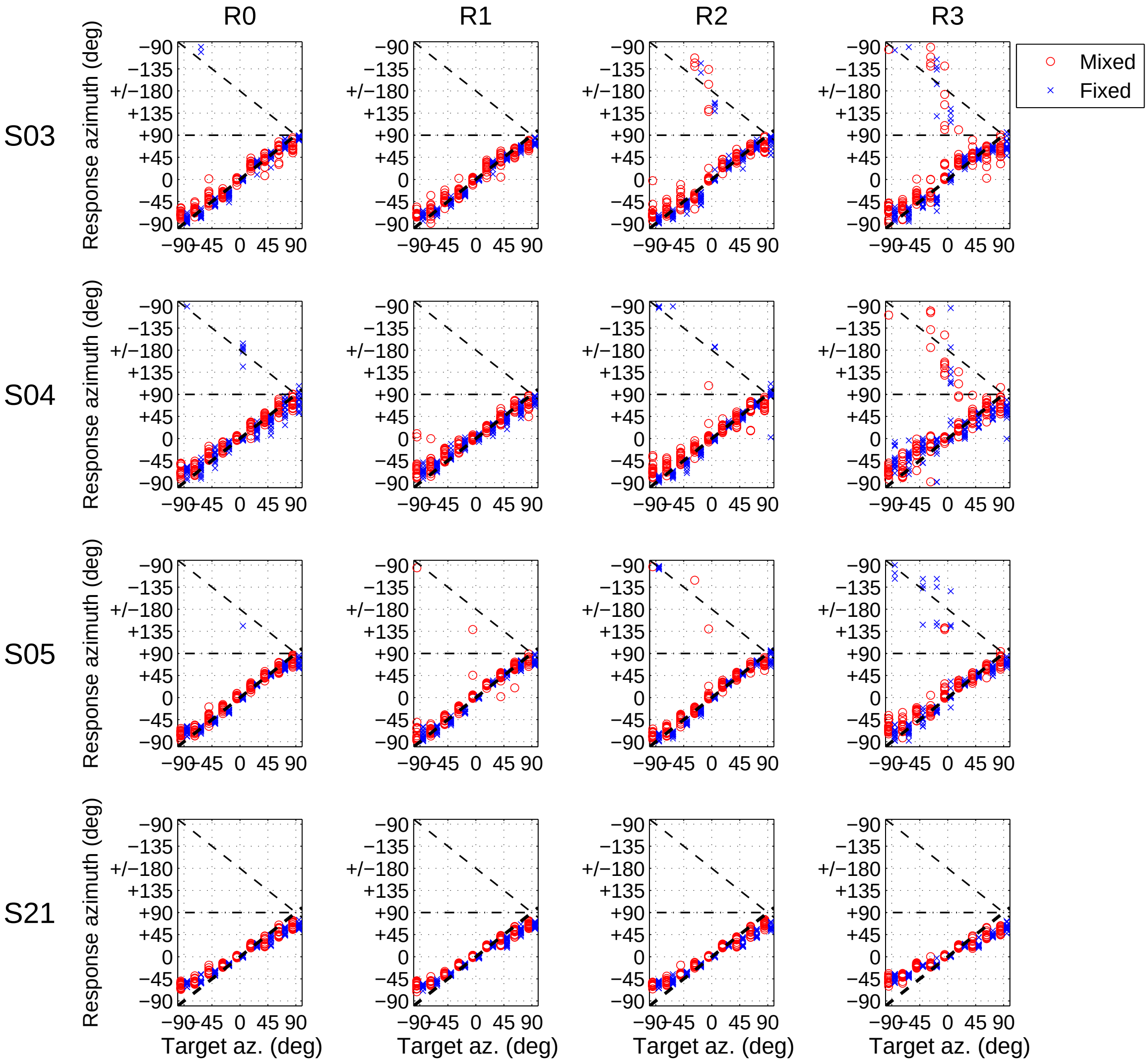


Figure 4.3: Localization responses from participants S03, S04, S05, and S32 in R0, R1, R2, and R3. Participants are organized in separate rows and rooms are in separate columns. x-axis coordinates were adjusted - slightly to the left for mixed trials (marked with red circles) and to the right for fixed trials (marked with blue crosses).

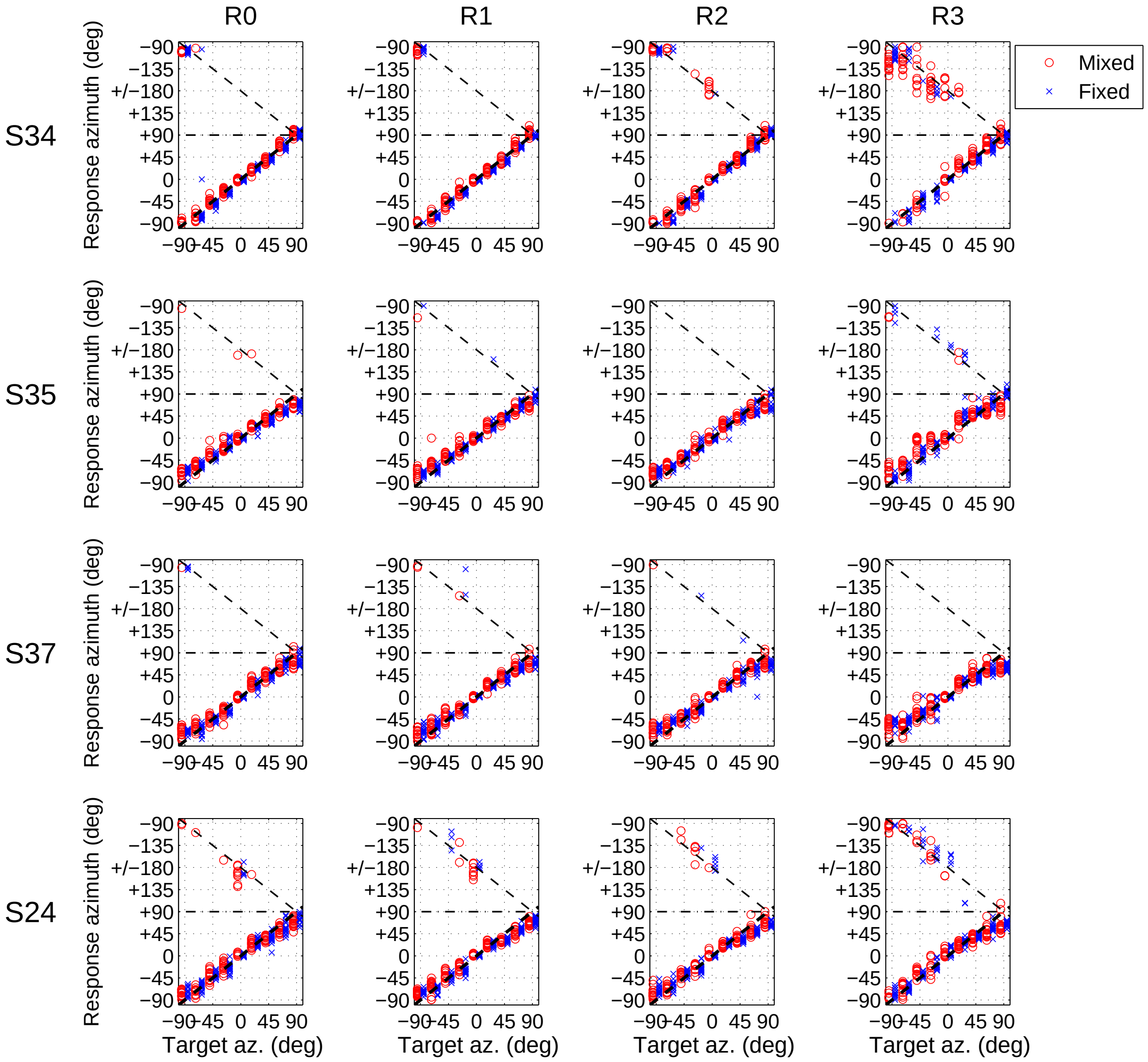


Figure 4.4: Similar to Fig. 4.3. Localization responses from participants S34, S35, S37, and S24 in R0, R1, R2, and R3.

ambiguous). The mean and standard deviation of the azimuths of the non-reversed responses across all individuals are plotted in Figs. 4.6 and 4.7.

A three-way repeated-measures analysis of variance (ANOVA) with room, type of

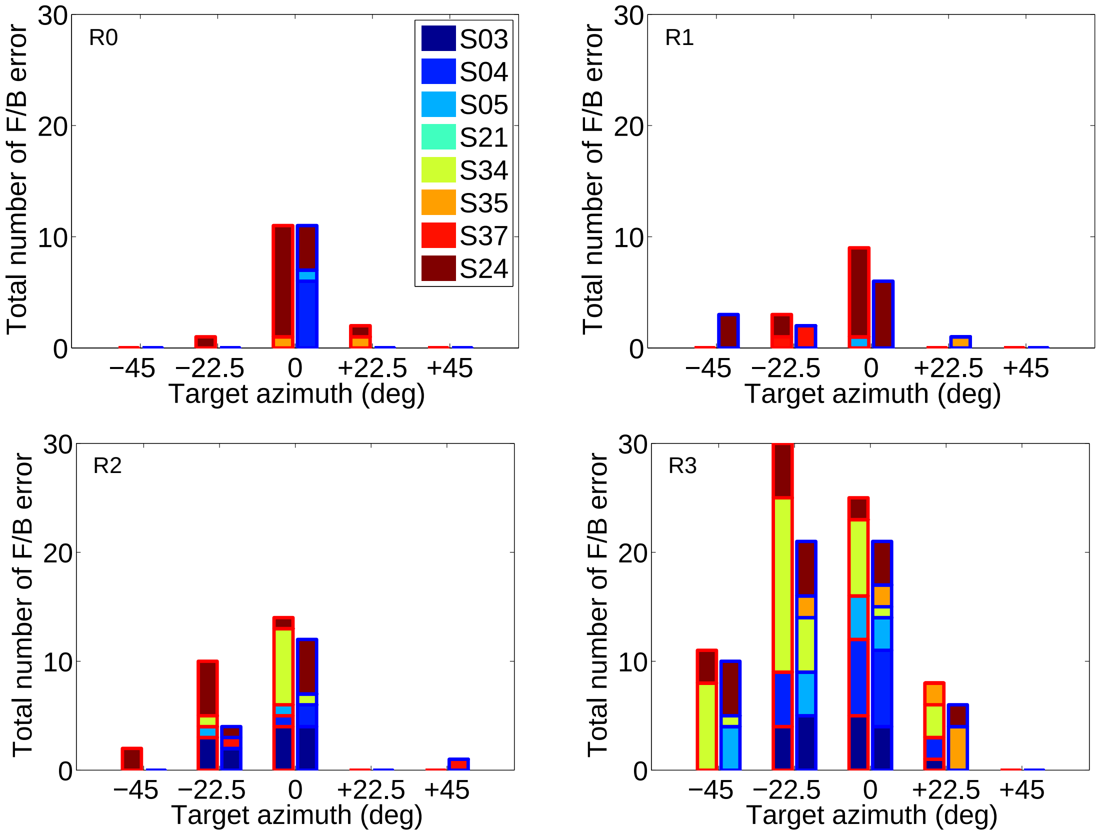


Figure 4.5: Total number of back-hemisphere responses to front-hemisphere targets (front/back errors) in R0 (top left panel), R1 (top right panel), R2 (bottom left panel), and R3 (bottom right panel). Total number of trials per condition was 16. Each colored section indicates the number of errors from each individual (legend in top-left panel). Lateral targets (absolute lateral angle > 45°) are excluded from the plot. For each pair of bars, left (red outline) and right (blue outline) bars correspond to mixed and fixed conditions, respectively.

exposure, and target location as factors ($4\times\ 2\ \times\ 9$) was applied on the mean non-reversed response azimuth data. The ANOVA table is shown in Table 4.1. The main effect of location was significant. None of the other factors or factor interactions yielded any statistically significant effect. Thus, we concluded that there was no systematic response bias in R1, R2, and R3 relative to the mean response in R0.

The standard deviation data (*cf.* Fig. 4.7) were averaged across locations, and

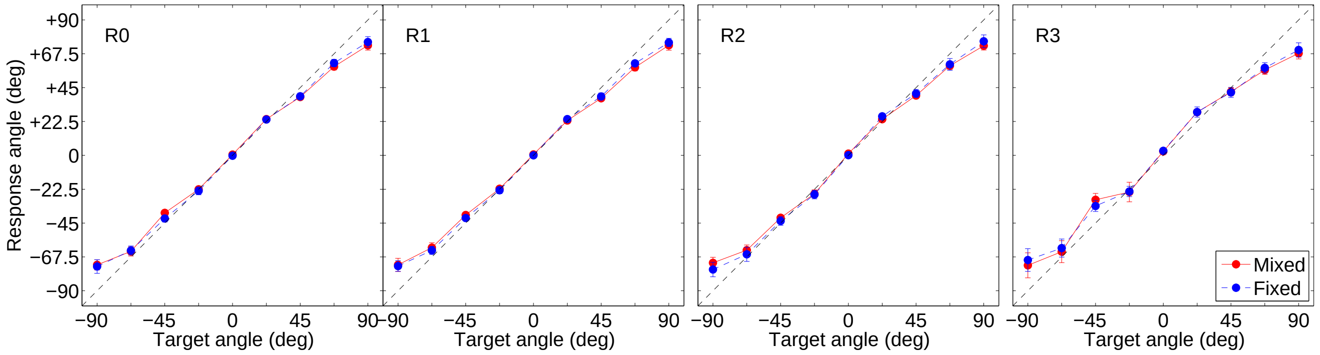


Figure 4.6: Mean azimuth of non-reversed responses across all individuals in R0, R1, R2, and R3 (left to right, respectively). Error bars indicate ±1 standard error of the mean.

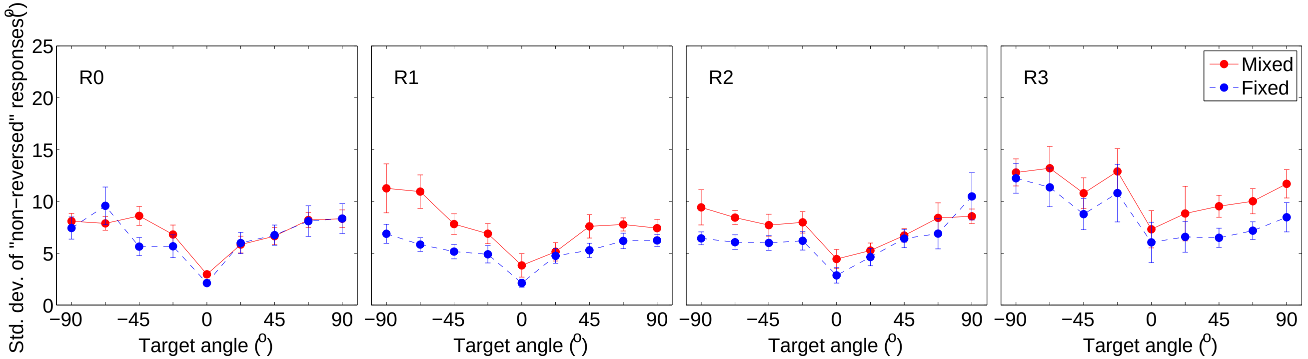


Figure 4.7: Standard deviation of azimuths of non-reversed responses averaged across all participants in R0, R1, R2, and R3 (left to right, respectively). Error bars indicate ±1 standard error of the mean.

Table 4.1: Results of the three-way repeated-measures ANOVA of the mean non-reversed response azimuth data.

Factor	F-, p-values, η^2	Signif.
Room	$p = 0.27$	
Exposure	$p = 0.95$	
Target az.	$F(1.380, 9.662) = 396.251, p < 0.001, \eta^2 = 0.98$	*
Room×Exposure×Target az.	$p = 0.77$	
Room×Exposure	$p = 0.65$	
Room×Target az.	$p = 0.12$	
Target az.×Exposure	$p = 0.34$	

the results are plotted in Fig. 4.8a. While no obvious bias was observed with the increase in lateral spread of the reflections, the variability of the azimuth responses was affected. The standard deviation of the azimuth responses was larger in R3 than in R0, R1, and R2. Averages of standard deviation in the mixed/fixed conditions were 7.04°/6.62° (0.49/0.76), 7.62°/5.26° (0.68/0.49), 7.43°/6.21° (0.65/0.54), and 10.79°/8.65° (1.15/1.17) in R0, R1, R2, and R3, respectively. Values inside the brackets are the standard errors. The R0-data agree with the results of another study (Yost, Loiselle, Dorman, Burns, & Brown, 2013) which found that the rms-error in sound localization by normally hearing listeners in the anechoic field was normally distributed with mean of 6.2° and a standard deviation of 1.79°.

A two-way repeated measures ANOVA with room and exposure type as factors (4×2) was performed on the across-location averaged standard deviation data. The ANOVA results are shown in Table 4.2a. The interaction between room and exposure type was significant. Two-tailed paired t-tests were applied to find whether the exposure effect was significant in each room condition. The results, as well as the average difference in standard deviation (mixed-fixed), are shown in Table 4.2b. Significant differences due to exposure were found in R1, R2, and R3.

Table 4.2: Two-way ANOVA and paired t-tests results for the standard deviation of non-reversed response data.

(a) Results of a two-way repeated-measures ANOVA for the standard deviation of non-reversed response azimuth data averaged across locations.

Factor	F-, p-values, η^2	Signif.
Room	$F(1.6, 11.3) = 13.018$, $p < 0.005$, $\eta^2 = 0.650$	*
Type of exposure	$F(1, 7) = 21.810$, $p < 0.005$, $\eta^2 = 0.757$	*
Room × type of exposure	$F(2.5, 17.6) = 3.575, p < 0.05, \eta^2 = 0.338$	*

(b) Results of paired t-tests (pairwise αs were adjusted with the Šidák correction factor) for the standard deviation of non-reversed response azimuth data in R0, R1, R2, and R3.

Room	$SD_{mixed} - SD_{fixed}$ (SE)	t-, p-values	Signif.
R0	0.42° (0.52°)	$p = 0.45$	
R1	2.37° (0.44°)	$t(7) = 5.340, p < 0.005$	*
R2	1.22° (0.50°)	$t(7) = 2.462, p < 0.05$	*
R3	2.14° (0.63°)	$t(7) = 3.418, p < 0.05$	*

While the effect of exposure in R1, R2, and R3 were not large, it seemed to be consistent across almost all listeners. Fig. 4.8b shows the individual data on the mixed-fixed differences in response variability. Almost all listeners achieved lower response variability in the fixed-exposure condition than in the mixed-exposure condition (a difference greater than zero in Fig. 4.8b).

4.4 Discussion

In this study, the effects of reverberation context and increased lateral spread of reflections on sound localization performance in reverberation were examined. The results indicate that listeners' mean responses were not statistically different across rooms, including R0. The absence of response bias, particularly in R3, seemed to contradict the steady-state interaural cues existing in the stimuli. Analyses of these cues in R3 (Sec. 3.6) indicated low interaural coherence (in the range of 0.2-0.6) in frequency bands ranging from 500-2000 Hz and high ITD variability across frequency bands. A bias towards the median plane can clearly be seen in the steady-state ILD cues. Thus, it seemed that the steady-state ILD cues had little effect on the localization responses. Possibly, listeners relied more on the ITD or onset cues. This is consistent with the well-known dominance of ITD for horizontal sound localization (Wightman & Kistler, 1992; Macpherson & Middlebrooks, 2002), as well as with the precedence effect, the dominance of onset cues when the ongoing spatial cues are ambiguous (Freyman et al., 1997).

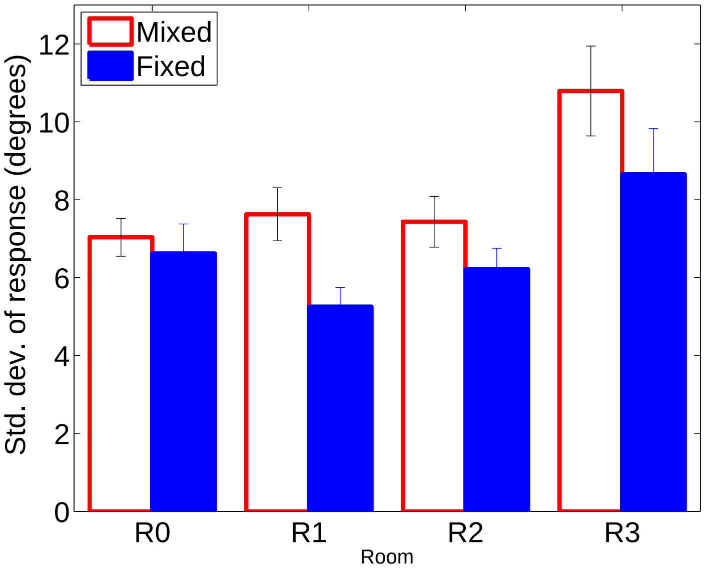
While increasing the lateral spread of reflections did not affect the mean azimuth responses, it increased the variability of the responses (Figs. 4.7 and 4.8a). Furthermore, a significant effect of exposure on the response variance was found in R1, R2, and R3. In these rooms, response variability was lower in the fixed-exposure condition than in the mixed-exposure condition. While the effect was not large, this pattern was rather consistent across listeners. Almost all, except one listener in R2, demonstrated this pattern. This result supports our hypothesis that sound localization performance depends not only on the stimulus in each trial, but also on the pattern in the overall stimulus set.

The underlying mechanisms contributing to the exposure effect, however, cannot be directly inferred from our study. The following mechanisms may occur. First, the variability in interaural coherence (IC) across time and frequency was in overall lower in the fixed-exposure condition than in the mixed-exposure condition (*cf.* Fig. 3.8). Results of a lateralization study by Le Goff et al. (2013) demonstrated that in the integration of ITD across frequency, the weight of the ITD in each frequency channel was proportional to its reliability (i.e., proportional to IC). The lower IC variability might have allowed listeners to weight ITD cues across frequency more optimally in the fixed-exposure condition than in the mixed-exposure condition. To test this hypothesis, future experiments utilizing stimuli that allow for quantification of cue weights could be performed to compare the cue weights under different reverberation contexts. Second, onset and ongoing cues might be weighted more optimally in the fixed exposure conditions. For example, in fixed-R1 condition, both onset and ongoing cues are equally reliable, since all reflections originated from the same location as the direct sound. In the mixed-exposure condition, however, R1-trials were mixed with R2- and R3-trials, which had less reliable ongoing cues. Possibly, listeners weighted the ongoing cue in R1-trials in the mixed exposure condition less heavily than in the fixed-R1 condition. Third, the consistency in reverberation might allow listeners to deconvolve the reverberation from the dry stimulus. However, this is an unlikely possibility; it has been shown that listeners cannot deconvolve their own anechoic HRTFs to identify stimuli with varying high-frequency spectral profile and roving locations along the median plane (Rakerd, Hartmann, & McCaskey, 1999; Macpherson, 1995).

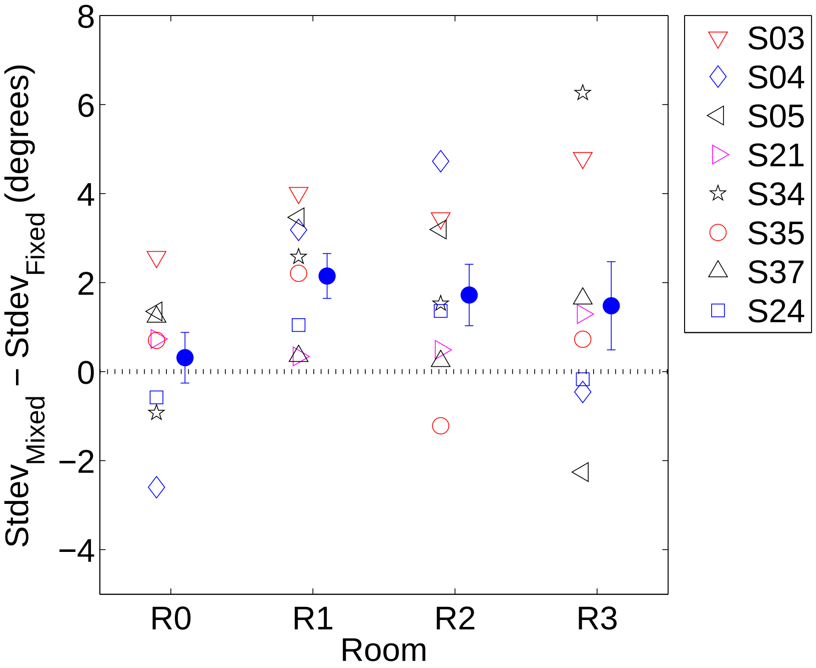
In studies examining generalization of learning in binaural advantage measured with word discrimination task (Nahum et al., 2010, see also the Introduction of this chapter) and in distance perception (Schoolmaster et al., 2003), it was found that the group of listeners who were trained with the consistent-exposure condition (*cf.* Fixed/Mixed group in our study) performed better than the participants in the other group (*cf.* Mixed/Fixed group) in both mixed- and fixed-exposure conditions. In the study by Nahum et al. (2010), the binaural advantage obtained in the mixed-exposure condition for participants trained with the fixed-exposure protocol was comparable to the binaural advantage obtained in the fixed-exposure condition. That is, listeners who were trained with the consistent protocol were able to identify and discriminate the $S_{\pi}N_0$ trials from the S_0N_0 trials and to fully utilize the IPD cue when the trials were later randomly interleaved. On the other hand, listeners who were trained with the mixed protocol could not fully utilize the IPD cue when such a cue was later made consistent. This suggests that prior exposure to a highly varying context may impede the ability to recognize a pattern and to utilize a cue when the stimuli are later presented in a consistent manner (i.e., when the uncertainty in the context is later reduced). To see whether our data exhibit a similar pattern, the individual standard deviation data are plotted in Fig. 4.9, identifying the participants in the Mixed/Fixed (empty markers) and Fixed/Mixed (filled markers) groups. Our data showed no indication that the Fixed/Mixed group performed better than the Mixed/Fixed group. A possible explanation of this result could be the smaller number of trials in our experiment (30 trials/location in each room and exposure combination). Nahum et al. (2010) trained their participants for more than 7 days while Schoolmaster et al. (2003) performed their experiment in 12 sessions. The duration of the fixed-exposure blocks in our experiment might have not been sufficient to observe a learning effect.

4.5 Summary

The result of this study provides support to the hypothesis that sound localization, particularly in reverberation, is influenced not only by the spatial information present in isolated trials, but also by the overall stimulus context. Presenting the stimuli in consistent reverberation context resulted in variability of azimuth responses that was lower than that obtained in the inconsistent reverberation context. Possibly, spatial cues were weighted differently according to the reverberation context, and the lower cue variability in the fixed condition allowed listeners to weight the cues across frequency and/or time more optimally. Future studies could be performed to determine the mechanisms underlying this effect and the time course and amount of improvement.



(a) Standard deviation of the non-reversed responses averaged across all locations.



(b) Individual differences between the standard deviation in the mixed and fixed conditions

Figure 4.8: Standard deviation of the non-reversed response in R0, R1, R2, and R3: (a) averaged across locations, (b) individual data, averaged across locations. Error bars indicate ± 1 standard error of the means.

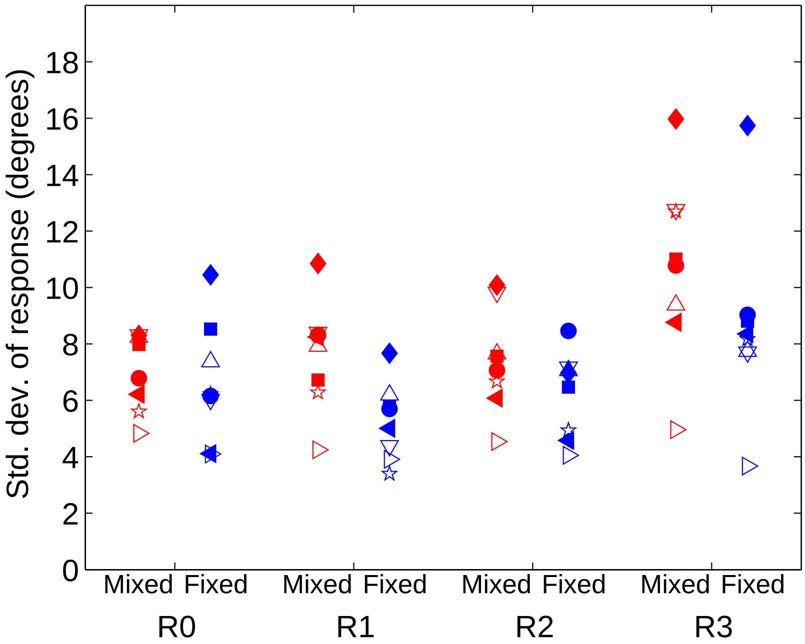


Figure 4.9: Individual location-averaged response variability (the corresponding means were plotted in Fig. 4.7). Empty markers are individuals in the Mixed/Fixed group. Filled markers are individuals in the Fixed/Mixed group.

Chapter 5

Experiment II: Effect of reverberation context on speech perception in noise

5.1 Introduction

The positive effect of consistent reverberation context on speech perception in a spatially separated noise has been demonstrated in several studies (Brandewie & Zahorik, 2010; Zahorik & Brandewie, 2011; Brandewie & Zahorik, 2013; Srinivasan & Zahorik, 2013) and discussed earlier in the Introduction chapter of this thesis (Section 1.1). Briefly, in those studies, speech intelligibility was tested with the target speech presented from the front (0°) and a Gaussian noise masker presented to the right of the listener (+90°). The task was performed under mixed- and fixed-exposure conditions; in the mixed-exposure condition, the reverberation (simulated room) was randomly varied across trials, whereas in the fixed-exposure condition, the simulated room was fixed across all trials in an experimental block. Results from these studies show that speech intelligibility performance was better under the fixed-exposure condition. The difference in performance relative to the mixed-exposure condition, measured with the CRM speech materials (Bolia et al., 2000), was approximately 2-3 dB lower in speech reception threshold (at 51% correct response rate), or approximately an 18% improvement in the identification of target words, for rooms with medium reverberation times, $0.3s < T_{60} < 1.2s$ (Brandewie & Zahorik, 2010; Zahorik & Brandewie, 2011).

Findings from related studies that examined the effect of reverberation context on a categorical "sir"-"stir" task (Watkins, 2005b; Watkins et al., 2011; Watkins & Raimond, 2013; Srinivasan & Zahorik, 2014, see also the Introduction, Chapter 1) suggest that consistent reverberation present in the context provides listeners with information about the reverberation tail, allowing them to compensate for the loss of modulation in the target words' amplitude envelope. However, besides the role of the context's amplitude envelope, there is little experimental evidence about other mechanisms that may contribute to the exposure effect found in these studies.

The study presented in this chapter aimed to examine whether a mechanism related to compensation of the interaural decorrelation of the target and masker signals plays any role in enhancing speech perception. The experiment employed a setup similar to that used in other studies (Brandewie & Zahorik, 2010; Zahorik & Brandewie, 2011; Srinivasan & Zahorik, 2013). The method used to simulate the reverberation in this experiment, however, was different from the latter studies which utilized BRTFs derived from a room model (Zahorik, 2009; Allen & Berkley, 1979). In the current study, the effects of reverberation context on speech perception in a spatially-separated noise were measured in the simulated rooms presented in Chapter 3 (R0, R1, R2, and R3). Recall that R1, R2, and R3 vary in the range of lateral spread of reflections (interaural coherence was highest in R1 and lowest in R3; see Fig. 3.8), but that their reverberation times are equal ($T_{60} \approx 0.47$ s); this reverberation time is similar to the broadband reverberation time of one of the rooms which yielded a 2.7-dB exposure effect in a previous study (Brandewie & Zahorik, 2010).

In other studies which used BRTFs derived from a room model, the set of reverberant environments employed varied in both their reverberation times and interaural coherence values, i.e., increase in reverberation time caused decrease in interaural coherence since the simulated rooms differed in their absorption coefficients and the source distance was fixed. By maintaining a similar reverberation time and varying only the lateral spread of the reflections in the set of reverberation patterns used in our experiment, we aim to obtain similar amplitude envelopes across the rooms. Thus, if an exposure effect is observed in our experiment, it may be attributed to consistency in interaural coherence.

In addition to stimuli in the preceding trials, the speech preceding the target words within a trial can also give information about the reverberation. In several studies (Brandewie & Zahorik, 2010; Zahorik & Brandewie, 2011), the effect of exposure was quantified as the difference in performance (speech reception threshold) between the mixed exposure condition, in which only the target words were presented ("colour number" in the CRM corpus) and the room was randomly varied across trials, and the fixed exposure condition, in which a sentence carrier preceded the target words ("Ready Baron go to colour number") and the room was fixed across trials. In these studies, it was shown that the presence of the sentence carrier did not affect intelligibility performance in the free field. However, in reverberation, the difference in performance between the mixed and fixed exposure conditions was significant, e.g., 2.7 dB in the study of Brandewie and Zahorik (2010). Thus, the exposure effect measured in these studies can be attributed to a combination of the effect of consistency of reverberation across trials and the effect of the sentence carrier.

In the present study, we also measured the effect of the reverberation context (i.e., exposure across trials) and the effect of the sentence carrier separately. The effect of the reverberation context was quantified as the difference in performance between the Mixed-Short and Fixed-Short conditions. The Mixed-Short condition was identical to the mixed exposure condition in previous studies (Brandewie & Zahorik, 2010; Zahorik & Brandewie, 2011), whereas in the Fixed-Short condition, only the target words were presented in each trial and the reverberation is fixed across trials. The effect of the sentence carrier was quantified as the difference in performance between the Fixed-Short and Fixed-Full conditions; the Fixed-Full condition, which used the full CRM sentences was comparable to the fixed exposure condition in the aforementioned studies.

5.2 Method

5.2.1 Participants

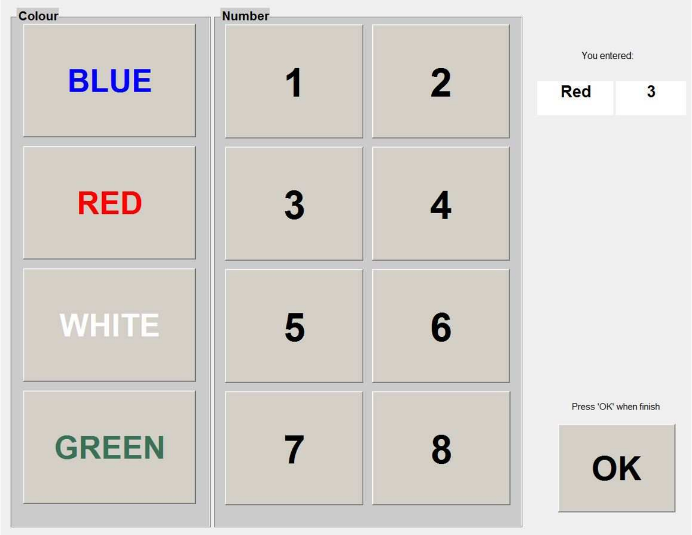
Ten young, normally hearing listeners who used English as their primary language participated in this experiment (7 females, 3 males; 19-22 years of age, mean age = 20.2). None of them participated in Experiment I. One participant (S18) also participated in Experiment IIIA. Total testing time per participant was about 4.5 hours, divided into 3 different-day sessions (∼1.5 hours each). Testing progressed according to individual pace and schedule. Short breaks were provided whenever needed.

5.2.2 Stimuli and procedure

Listeners were presented with either a short CRM sentence in the form of "$\langle$*colour*$\rangle$ $\langle$*number*$\rangle$" or a full CRM sentence in the form of "Ready Charlie go to $\langle$*colour*$\rangle$ $\langle$*number*$\rangle$ now". The short and full versions of the CRM sentences have been discussed in more detail in Sec. 3.4. Furthermore, the full CRM sentences were scaled so that the RMS-levels of their colour and number portions matched the RMS-levels of the short CRM samples.

Each short or full CRM sentence was presented with a simultaneous constant-amplitude speech-shaped noise. To create the speech-shaped noise, a Gaussian noise was filtered with a 256-point finite impulse response filter with a spectrum magnitude matching the long-term average spectrum magnitude of the short CRM sentences. The latter was estimated by averaging the spectrum magnitude of 40-ms Hann-windowed segments (overlapping by 50%) of all short CRM samples. A new instance of the speech-shaped noise was generated for each trial. In each trial, the speech was presented from 0° and the noise was presented from +90°. The onset of the noise preceded that of the speech by 1 second, and the noise remained on for 1 second after the end of the speech. Both the target and noise signals were reverberated according to the method described in Sec. 3.5. The noise level was constant at 65 dB SPL, and the target speech was presented at SNR levels of -6, -10, -14, and -18 dB.

Participants were seated at the centre of the loudspeaker array (see Fig. 3.1) in the anechoic chamber. They were instructed to face the 0°-loudspeaker and to stay still during each trial. An experimenter monitored their position, and reminded the participants to correct their positions if needed.



You entered:

Red 3

Figure 5.1: Graphical user interface used in the CRM speech perception task.

Participants were asked to identify the colour and number in a presented sentence, and entered their responses via an 8” LCD touch-screen monitor (Accelevision). The interface is shown in Fig. 5.1. Upon pressing the ‘OK’ button, the next trial was presented after a brief delay (approximately 1 second). Listeners received no feedback regarding their responses. They were informed of the type of sentences to be presented (short or full-length sentences) at the beginning of each block.

Practice trials were presented at the beginning of the experiment, consisting of 10 short sentences and 10 full-length CRM sentences in R0 at an SNR of -6 dB. This was an easy condition with high target audibility; all listeners were able to quickly learn the procedure within a few trials.

Listening conditions consisted of four simulated rooms (R0, R1, R2, and R3) and three types of exposure: Mixed-Short, Fixed-Short, Fixed-Full. In the Mixed-Short condition, the rooms were varied in a random order across trials, and a short CRM sentence was presented on each trial. In the Fixed-Short condition, the room was fixed across trials, and a short CRM sentence was presented on each trial. In the Fixed-Full condition, the room was fixed across trials, and a full-length CRM sentence was presented on each trial. In each of the listening conditions, speech stimuli were presented at 4 SNRs with 30 trials for each SNR. The SNR on each trial was randomized. There were 1440 trials in total which were divided into 8 blocks of Mixed-Short trials, 8 blocks of Fixed-Short trials, and 8 blocks of Fixed-Full trials. Each block consisted of 60 trials, and the order of the blocks was randomized. However, for the Fixed-Short and Fixed-Full conditions, the two blocks corresponding to the same room were consecutively tested within the same session.

The CRM samples were randomly chosen (with replacement) with a balanced number of stimuli from each of the 8 talkers in the corpus.

5.3 Results

A trial was scored as correct if both the colour and number were correctly identified. The across-listener average proportion of correct responses for each listening condition is plotted in Fig. 5.2.

Speech intelligibility performance in each listening condition was quantified as the SNR corresponding to a fixed percentage (71%) of correct responses; let us define this measure as $\mathrm{SNR}_{71\%}$. An estimate of $\mathrm{SNR}_{71\%}$ was obtained for each listener and listening condition combination via its psychometric function, $P(x)$. $P(x)$ was estimated from 4 data points (proportion of correct response at -6, -10, -14, and -18 dB) by fitting a logistic function of the following form,

$$P(x) = \gamma + \frac{1-\gamma}{1+e^{-\beta(x-\alpha)}}. \tag{5.1}$$

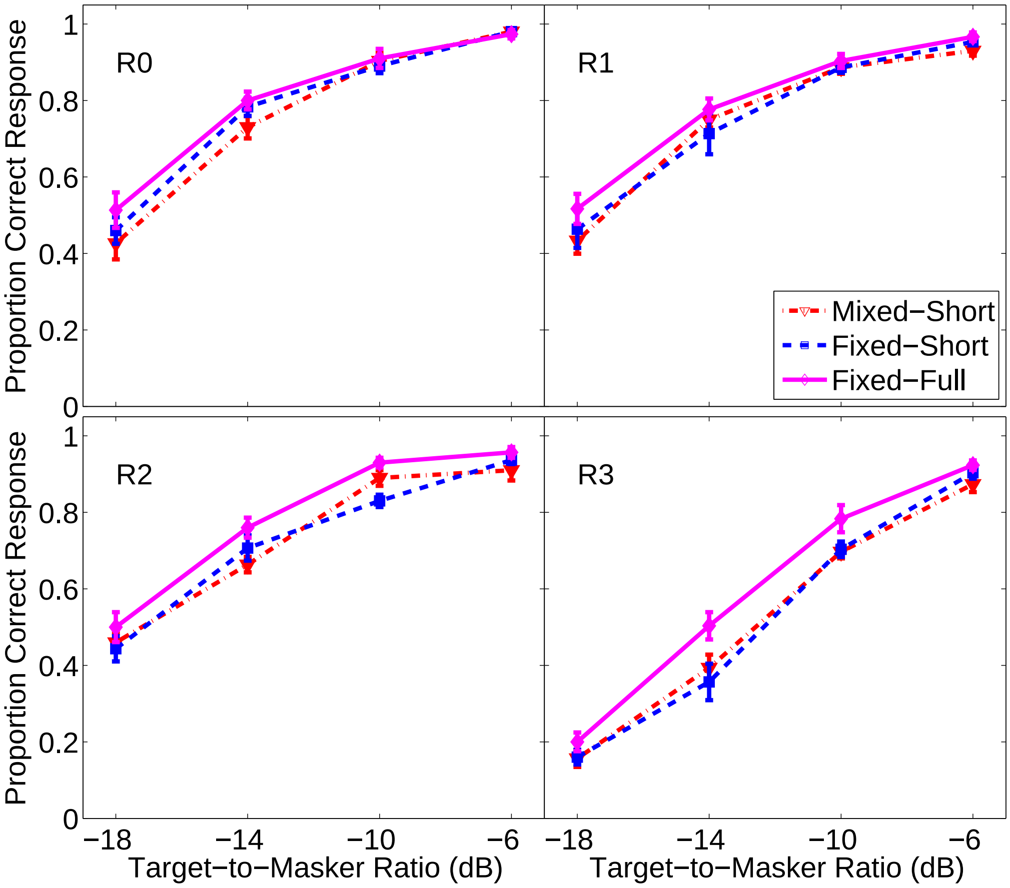


Figure 5.2: Performance in the speech-in-noise task in R0 (top-left panel), R1 (top-right panel), R2 (bottom-left panel), and R3 (bottom-right panel) averaged across 10 participants. Performance for different sentence and exposure types are plotted within each panel. Error bars indicate ±1 standard error of the mean.

$P(x)$ is the proportion of correct trials at x-dB SNR. γ was set to a constant 1/32 (i.e., the chance performance of this task, given 32 possible colour and number combinations). α and β are the midpoint and slope parameters, respectively. They were estimated using MATLAB's `nlinfit` routine, and their 95% confidence intervals (CI) were obtained with the `nlparci` routine. The estimated parameters, 95% CIs, and goodness-of-fit (R^2) data are included in Appendix B.

A plot of the individual $\text{SNR}_{71\%}$ data can be seen in Fig. 5.3, and the corresponding means and standard deviations are shown in Table 5.1. $\text{SNR}_{71\%}$ increased with increasing lateral spread of reflections. Averaged across listening conditions, the $\text{SNR}_{71\%}$s in R0, R1, R2, and R3 were -14.9 dB (SE = 0.40), -14.8 dB (SE = 0.48), -14.1 dB (SE = 0.31), and -10.3 dB (SE = 0.18), respectively. This decline in performance was expected from a loss of both better-ear advantage and binaural benefit.

Table 5.1: The across-listener mean and standard deviation values of $\text{SNR}_{71\%}$ in R0, R1, R2, and R3.

Room	$\text{SNR}_{71\%}$ (dB)					
	Mixed-Short		Fixed-Short		Fixed-Full	
	Mean	SD	Mean	SD	Mean	SD
R0	-14.46	1.51	-14.91	1.15	-15.44	1.70
R1	-14.30	1.33	-14.48	2.20	-15.48	1.84
R2	-13.49	0.99	-13.66	1.26	-15.18	1.12
R3	-9.69	0.69	-9.99	0.65	-11.29	1.39

The $\text{SNR}_{71\%}$ data were submitted to a two-way repeated measures ANOVA with room (R0, R1, R2, R3) and type of exposure (Mixed-Short, Fixed-Short, Fixed-Full) as factors. The results indicated that the interaction between room and type of exposure was not statistically significant (p = 0.667). Both the main effects of

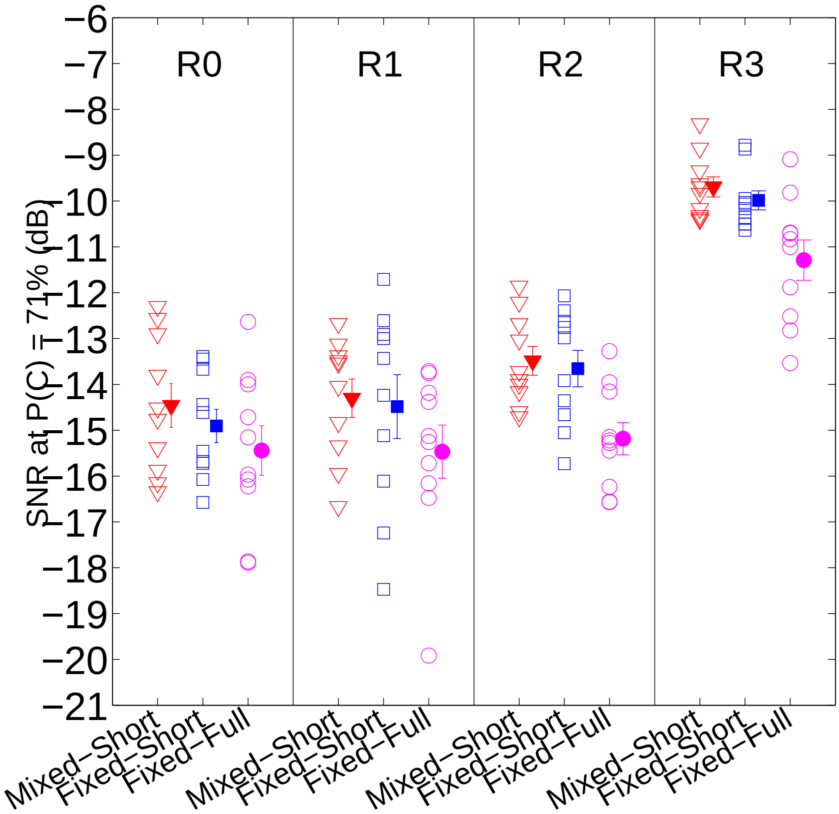


Figure 5.3: The SNR at 71% correct response rate ($\mathrm{SNR}_{71\%}$) for each participant in R0, R1, R2, and R3 (left to right panel, respectively). Filled symbols are average data. Error bars indicate ±1 standard error of the mean. Data were estimated from the fitted logistic functions in eq. (5.1).

room and type of exposure were statistically significant [Room: $F(1.865, 16.782) = 101.113, p < 0.001, \eta^2 = 0.918$; Type of exposure: $F(1.694, 15.243) = 21.606, p < 0.001, \eta^2 = 0.706$].

As expected, $\mathrm{SNR}_{71\%}$ increased (i.e., performance decreased) with an increase in the lateral spread of reflections. This was likely due to reduced better-ear and binaural squelch advantages. Post-hoc pairwise comparisons (αs adjusted with the Šidák correction factor) indicated statistically significant differences in SNR$_{71\%}$ between R0 and R2 ($p < 0.05$), R2 and R3 ($p < 0.001$), R1 and R3 ($p < 0.001$), and R0 and R3 ($p < 0.001$).

Averaged across all rooms, SNR$_{71\%}$s in the Mixed-Short, Fixed-Short, and Fixed-Full conditions were -13.0 dB (SE = 0.37), -13.3 dB (SE = 0.32), and -14.3 dB (SE = 0.31), respectively. Post-hoc pairwise comparisons (αs adjusted with the Šidák correction factor) indicated that the difference between SNR$_{71\%}$s in the Mixed-Short and Fixed-Short conditions (the effect of reverberation context) was not statistically significant ($p = 0.632$). Statistically significant differences were found between SNR$_{71\%}$s in the Fixed-Short and Fixed-Full conditions (the effect of sentence carrier; $p < 0.01$) and between SNR$_{71\%}$s in the Mixed-Short and Fixed-Full conditions (the combined effect of sentence carrier and consistent reverberation; $p < 0.001$).

Individual pairwise performance differences between different types of exposure in each room are plotted in Fig. 5.4. Data points plotted with blue triangles indicate the effect of reverberation context, a positive value indicates the benefit of consistent reverberation. Data points plotted with black asterisks indicate the effect of sentence carrier; a positive value indicates the benefit of sentence carrier. Data points plotted with red triangles are the combined effect of sentence carrier and consistent reverberation. The average differences in SNR$_{71\%}$s between the Fixed-Short and Fixed-Full conditions were 0.54 dB, 0.98 dB, 1.53 dB, and 1.30 dB in R0, R1, R2, and R3, respectively. The average differences in SNR$_{71\%}$ between the Mixed-Short and Fixed-Full conditions were 0.98 dB, 1.18 dB, 1.69 dB, and 1.60 dB in R0, R1, R2, and R3, respectively. The differences between the Fixed-Short and Fixed-Full thresholds, and between the Mixed-Short and Fixed-Full thresholds were statistically

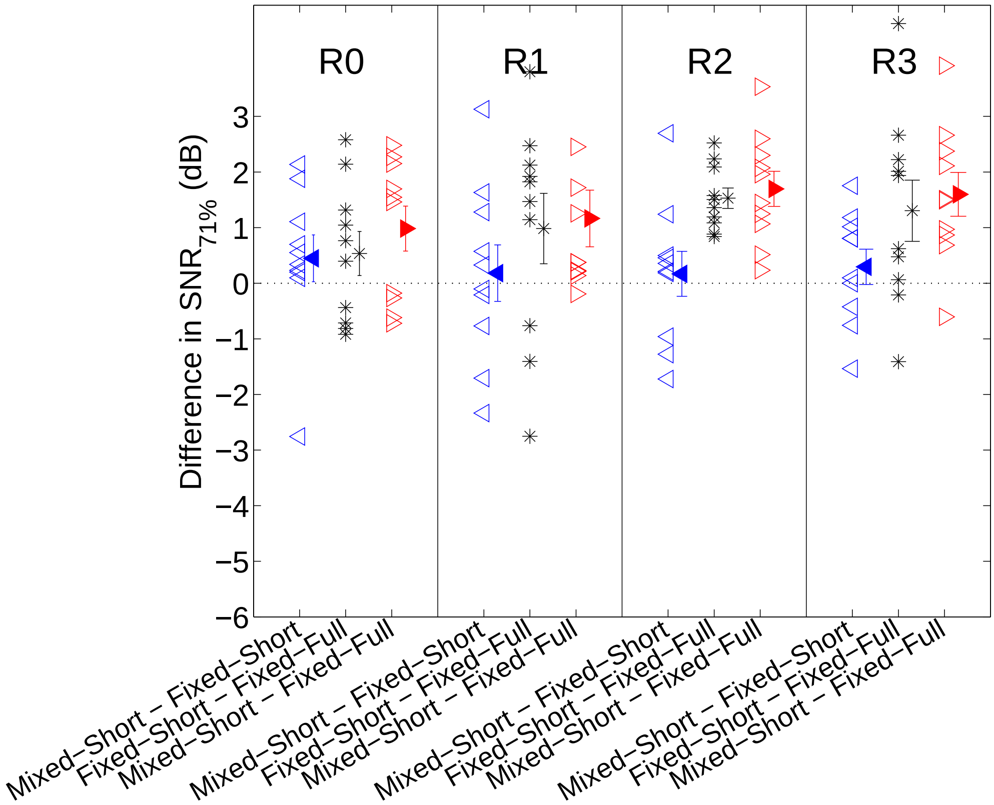


Figure 5.4: Differences in $\mathrm{SNR}_{71\%}$ between listening conditions for each participant in R0, R1, R2, and R3 (left to right panel, respectively). Mixed-Short - Fixed-Short indicates the effect of the type of exposure. Fixed-Short - Fixed-Full indicates the effect of the sentence carrier. Mixed-Short - Fixed-Full indicates the compound effect of exposure and sentence carrier. Filled symbols are the means, and error bars indicate ±1 standard errors of the mean.

different from 0. However, the interaction between room and type of exposure was not significant, indicating that the differences due to the type of exposure were not statistically different across R0, R1, R2, and R3.

5.4 Discussion

In contrast to Experiment I, in this task listeners did not seem to benefit from the consistency of reverberation across trials; the $\text{SNR}_{71\%}$s obtained in the Mixed-Short and Fixed-Short conditions were similar. Additionally, the amount of improvement in speech intelligibility attributed to a reverberant sentence carrier in R1, R2, and R3 was much less than that observed in other studies (Brandewie & Zahorik, 2010; Zahorik & Brandewie, 2011). In those studies, the effect of the sentence carrier in reverberant listening environments was significantly larger than that obtained in the free field. In our experiment, the amount of improvement that was attributed to the sentence carrier in R1, R2, and R3 was on average larger than that obtained in R0 (free field); however, ANOVA results indicated that the amount of improvement in R1, R2, and R3 was not statistically different from the amount of improvement in R0.

The similarity in reverberation times among R1-, R2-, and R3-reverberation could contribute to the lack of exposure effect in our study. This would also indicate that consistency in reverberation did not yield additional compensation for degradation in interaural coherence. Perhaps the exposure effect obtained in the previous related studies (Brandewie & Zahorik, 2010; Zahorik & Brandewie, 2011; Srinivasan & Zahorik, 2013) was mainly attributed to the amplitude envelope in the consistent reverberation context rather than to a consistency in interaural coherence. This would be consistent with findings from other studies that investigated the roles of the amplitude envelope and fine-structure cues in the exposure effect (Watkins et al., 2011; Srinivasan & Zahorik, 2014). Findings of those studies suggested that the exposure effect could be mainly attributed to the context's amplitude envelope.

In our experiment the sentence carrier provided on average a 0.54-dB advantage in intelligibility in R0. This amount is comparable to the 0.84-dB advantage obtained by Brandewie and Zahorik (2010). The advantages of the sentence carrier in R1, R2, and R3 obtained in our experiment were 0.98, 1.53, and 1.30 dB in R1, R2, and R3, respectively, considerably smaller than the 2.7-dB advantage obtained by Brandewie and Zahorik (2010) in a room that has similar T_{60} and C_{50} as those used in the current study. Additionally, our data exhibited a considerable amount of inter-individual variability. The small effect sizes and individual variability in our data would have made it difficult to detect differences of effect sizes across room conditions. The data from the study of Zahorik and Brandewie (2011) indicated that the exposure effect was dependent on the reverberation time, and that the effect size seemed to improve with reverberation time, provided that $0.3s \leq T_{60} \leq 1.2s$. Future studies might measure the effect of coherence context in rooms with reverberation times longer than that used in the present study ($T_{60} \approx 0.47$ s).

The differences between the reverberation characteristics used in the present study and other studies could have also contributed to the discrepancy in findings. In related studies (Brandewie & Zahorik, 2010; Zahorik & Brandewie, 2011; Srinivasan & Zahorik, 2013, 2014; Watkins, 2005b; Watkins et al., 2011), the experiments were conducted under headphones, employing BRTFs measured in real rooms or derived via a room model that simulated real-room reverberation characteristics more closely than the simulated reverberation used in our study. For example, the simulated reverberation used in the present study did not have the lowpass characteristics typically found in real-room reverberation. Additionally, the late part of the simulated reverberation used in our study was not diffuse except in R3. Possibly, the exposure effect found in other studies was related to aspects of reverberation that were not captured in our simulated reverberation, thus contributing to the failure to find a significant exposure effect in our study.

In summary, the results of the current experiment indicated that consistency in coherence context did not yield a significant advantage in the speech intelligibility task. This suggests that consistency in interaural coherence does not significantly contribute to the effect of reverberation context found in similar experiments performed by other researchers. However, further investigations employing more realistic reverberation and longer reverberation times need to be performed to support this conclusion. Further investigations on the mechanisms underlying the exposure effect in this task may also look at the role of the amplitude modulation in the stimulus context.

Chapter 6

Experiment III: Effect of reverberation context in multi-talker listening situations

In the previous studies described in this thesis, listeners' abilities to adapt to reverberation were examined in single-source localization (Experiment I) and speech-in-noise tasks (Experiment II). In those tasks, there was little ambiguity about the identity of the target source (i.e., little informational masking). The studies presented in this chapter address the effects of reverberation context in more complex listening situations in which the target and maskers are all speech spoken by same-sex talkers. The benefit of talker separation on target intelligibility (i.e., the spatial release from masking) and the ability to identify the location of a target speech token within a multi-talker mixture were examined under consistent and inconsistent exposures to reverberation.

6.1 Introduction

The relationship between spatial hearing and speech intelligibility in multi-talker listening situations has been addressed in the Background chapter of this thesis

(Sec. 2.4.3). Briefly, spatial separation of talkers provides sequential grouping cues which facilitate the perceptual segregation of a speech mixture into distinct speech streams (Darwin & Hukin, 1997, 1998). Knowledge of the target location and continuity in the target location also allow listeners to selectively attend to a speech stream based on its location (Kidd, Arbogast, et al., 2005; Best et al., 2008). Both of these mechanisms enhance the identity of the target speech, thus reducing the effect of informational masking.

In Experiment I, we compared listeners' abilities to localize a single source in simulated reverberation environments in mixed and fixed reverberation contexts, and the results indicated a smaller rms-error in the azimuth judgements in the fixed-exposure condition. The current studies were designed to investigate whether the benefit of consistent reverberation context, as found in the single-source localization task, can also be obtained for speech intelligibility and localization tasks in multi-talker listening situations.

In Experiment IIIA (speech identification task), listeners' ability to identify the keywords in a target sentence presented simultaneously with two competing sentences was evaluated under mixed and fixed exposures to the reverberation. Speech identification performance was measured in co-located (talkers at 0°, 0°, 0°) and spatially-separated (talkers at -22.5°, 0°, +22.5°) talker configurations (Fig. 6.1). In both configurations, the target talker was always positioned at 0°. Listeners were aware of the target position and of the spatial configuration of the talkers. The benefit of talker separation (i.e., the spatial release from masking) was measured by comparing listeners' speech identification performance in the co-located and spatially-separated conditions.

Experiment IIIB (location identification task) examined whether the exposure effect obtained in Experiment IIIA could be attributed to improved localizability of the target talker. In Experiment IIIB, listeners' ability to identify the location of a target talker within a multi-talker configuration was evaluated in the mixed and fixed reverberation contexts. Three same-sex talkers were positioned at -22.5°, 0°, or +22.5°, and the probability of the target being located at one of the three possible directions (-22.5°, 0°, or +22.5°) was equally likely. This task is analogous to a listening situation in which listeners have no prior knowledge of the target talker's location. For example, in a group conversation, the target talker may change from time to time, thus requiring the listeners to locate and direct their attention to the new target.

6.2 Experiment IIIA: The effect of consistent reverberation on multi-talker speech perception

6.2.1 Method

Participants

Eleven young, normally hearing listeners who used English as their primary language participated in this experiment (2 males, 9 females; 20-27 years of age, mean age = 24). One participant (S03) participated in Experiment I, and one participant (S18) participated in Experiment II. Total testing time for each participant was 5 to 6 hours including short breaks. Participants typically completed the experiment in 3 different-day sessions, each session was approximately 2 hours. Testing progressed according to individual pace and schedule. Short breaks were provided whenever needed.

Stimuli and procedure

Sentences from the Coordinate Response Measure (CRM) corpus (Bolia et al., 2000) were used as stimuli. In each trial, participants were presented with three simultaneous CRM sentences. The sentences were from different but same-sex talkers (i.e., all-male or all-female talkers in each trial). The callsigns, colours, numbers, and same-sex talkers were chosen randomly without replacement for each trial with the restriction that the target callsign was always "Baron". The number of trials with male and female talkers was equal for each listening condition.

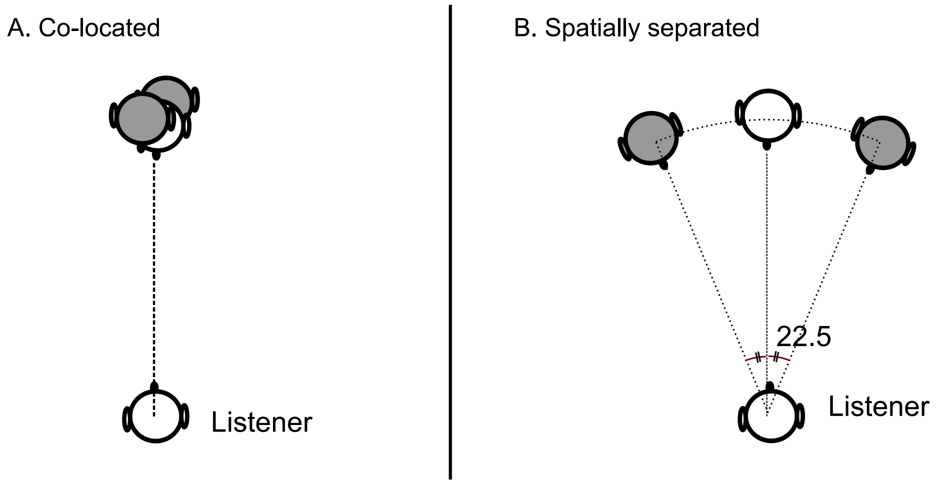


Figure 6.1: Spatial configurations of the talkers in: A) co-located condition and B) spatially-separated condition. The target talker (illustrated with an uncoloured cartoon head) was always positioned at 0°.

The sentences were presented in two spatial configurations as shown in Fig. 6.1, namely, co-located or spatially separated. The target sentence always originated from 0°. In the co-located condition, both competing sentences also originated from 0°. In the spatially-separated condition, one of the competing sentences was positioned at -22.5°, and the other was at +22.5°.

Reverberated stimuli (R1-, R2-, and R3-stimuli) were generated by processing each of the dry CRM sentences according to the method described in Sec. 3.5. Additionally, there were two types of listening exposure conditions: mixed and fixed. In the mixed-exposure condition, the simulated room was randomly varied across trials, whereas in the fixed-exposure condition, the simulated room was constant across trials. Participants were tested in their ability to identify the colour and number in the target sentence. They were asked to enter their responses via an interface displayed on an 8" touch-screen monitor (Accelevision). The interface was identical to that shown in Fig. 5.1. Both colour and number responses had to be provided in each trial, and the participants were instructed to provide their best guess or choose a random colour and/or number when they were not sure about the answers. No feedback on the correct answers was provided. They were instructed to face the front loudspeaker and to stay still during each trial. No physical restraint was used to maintain listeners' positions, but the experimenter monitored listeners' head positions and reminded them to maintain their head orientations from time to time.

At the beginning of the experiment, participants were given written instructions and some examples (with high target audibility) to familiarize them with the stimuli and the task procedure. They were also given practice trials consisting of 30 R0-trials at a target-to-masker ratio (T/M) of 4 dB in both the co-located and spatially separated conditions to examine whether they could take advantage of the talker separation. The T/M is defined as the ratio (in dB) of the level of the target sentence to the individual levels of the competing sentences. Both competing sentences were presented at the same level.

There were 16 listening conditions in total, consisting of 2 types of listening exposure (mixed and fixed), 2 types of talker configuration (co-located and spatially-separated), and 4 rooms (R0, R1, R2, and R3). In each listening condition, listeners were tested at 6 values of T/M. T/Ms were presented in decreasing order either block-by-block (mixed exposure condition) or trial-by-trial within a block (fixed exposure condition). The level of the target sentence was fixed at 60 dB SPL, while the levels of the competing sentences were adjusted according to the T/M. For R0- and R1-stimuli, the T/Ms tested were -4, -2, 0, 2, 4, and 6 dB for the spatially separated condition, and -2, 0, 2, 4, 6, 8 dB for the co-located condition. For R2- and R3-stimuli, the T/Ms tested were -2, 0, 2, 4, 6, and 8 dB for the spatially-separated condition, and 0, 2, 4, 6, 8, and 10 dB for the co-located condition. These particular T/Ms were chosen based on the results of a pilot study which was conducted to examine the range of T/M that would provide enough data points that span most listeners' psychometric functions. For each T/M, 30 trials were presented.

Trials were organized into experimental blocks as shown in Table 6.1. There were 12 blocks (6 co-located and 6 spatially separated) in the mixed exposure condition (Table 6.1a) and 8 blocks (4 rooms×2 talker configurations) in the fixed exposure condition (Table 6.1b). Each row in the tables corresponds to an experimental block. In the mixed-exposure condition, all blocks corresponding to a particular talker configuration were presented consecutively in the order of decreasing T/M. Trials in the fixed-exposure condition were also blocked by the room and talker configuration (e.g., R0/co-located and R0/spatially separated blocks were presented consecutively), and the trials within each room/talker-configuration block were presented in the order of decreasing T/M. For the fixed exposure condition, the room order was randomized

Table 6.1: The organization of trials into experimental blocks in: a) mixed-exposure condition, b) fixed-exposure condition.

(a) The grouping of trials in the mixed-exposure condition. An experimental block corresponds to a row in a table (120 trials/block). Experimental blocks were presented in the order of highest to lowest T/M.

Co-located				
Test order	T/M (dB)			
	R0	R1	R2	R3
1	8	8	10	10
2	6	6	8	8
3	4	4	6	6
4	2	2	4	4
5	0	0	2	2
6	-2	-2	0	0

Spatially separated				
Test order	T/M (dB)			
	R0	R1	R2	R3
1	6	6	8	8
2	4	4	6	6
3	2	2	4	4
4	0	0	2	2
5	-2	-2	0	0
6	-4	-4	-2	-2

(b) The grouping of trials in the fixed-exposure condition. Each row in a table corresponds to an experimental block (180 trials/block). Trials were blocked by the room and talker configuration. Trials within a room/talker-configuration block were presented in the order of the highest to lowest T/M.

R0	
Talker config.	T/M (dB)
Co-located	8, 6, 4, 2, 0, -2
Spatially separated	6, 4, 2, 0, -2, -4

R1	
Talker config.	T/M (dB)
Co-located	8, 6, 4, 2, 0, -2
Spatially separated	6, 4, 2, 0, -2, -4

R2	
Spatial config.	T/M (dB)
Co-located	10, 8, 6, 4, 2, 0
Spatially separated	8, 6, 4, 2, 0, -2

R3	
Spatial config.	T/M (dB)
Co-located	10, 8, 6, 4, 2, 0
Spatially separated	8, 6, 4, 2, 0, -2

Six participants first completed the mixed-exposure condition followed by the fixed-exposure condition. The other five participants performed the test in the reverse order. At the beginning of each block, information about the talker configuration for the particular block, as well as the target callsign ("Baron"), was displayed on the screen. Participants were informed and occasionally reminded that the target sentence was always presented from the front. They were also informed that the testing was to progress from easy to difficult target audibility.

6.2.2 Results

A trial was scored as correct if both the target colour and number were correctly identified. The percentages of correct response averaged across listeners in all listening conditions and T/M combinations are shown in Fig. 6.2.

Speech intelligibility performance in a listening condition was quantified as the T/M corresponding to a fixed correct-response rate. This was obtained via the estimated psychometric function. For each listener and listening condition, a psychometric function was estimated from 6 data points, each corresponded to performance at a particular T/M, by fitting those points with a logistic function of the following form,

$$P(x) = \gamma + \frac{1-\gamma}{1+e^{-\beta(x-\alpha)}},$$

where $P(x)$, $0 \leq P(x) \leq 1$, is the proportion of correct response at $T/M = x$ dB. γ was fixed at 1/32 (i.e., chance performance with 32 possible colour and number combinations). α and β are the threshold and slope parameters, respectively. The threshold, α, corresponds to the T/M which yields 51.5% correct performance rate. Estimates of α and β were obtained with the `nlinfit` routine in MATLAB. The

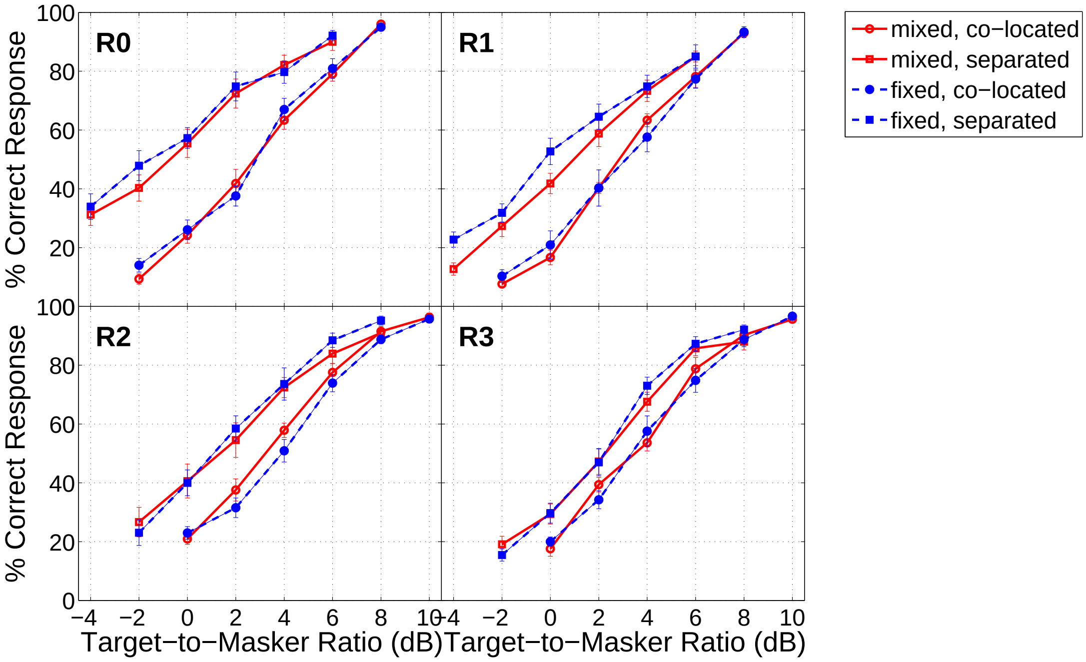


Figure 6.2: The average performance across all listeners in the multi-talker speech perception task. Error bars indicate ±1 standard error of the mean.

95% confidence intervals (CIs) of the estimated parameters were obtained with the `nlparci` routine. The estimated parameters, 95% CIs, and their goodness-of-fit (R^2) data are included in Appendix C.

The average threshold across listeners in each listening condition is plotted in Fig. 6.3. As expected, co-located thresholds were higher than spatially-separated thresholds, indicating a positive benefit of the talker separation in this task. Threshold data were submitted to a three-way repeated-measures ANOVA with room, talker configuration, and type of exposure as factors (4×2×2). The results are summarized in Table 6.2.

The three-way interaction of room, talker configuration, and type of exposure in

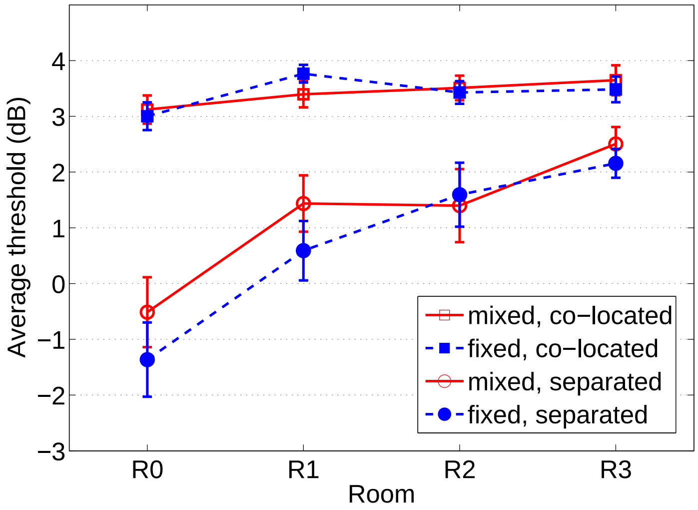


Figure 6.3: Speech reception thresholds, α, for the multi-talker speech perception task in all listening conditions averaged across all listeners ($n = 11$). Error bars indicate ±1 standard error of the mean.

Table 6.2: Results of the three-way repeated-measures ANOVA on the thresholds for the multi-talker speech identification task.

Factor	F-, p-values, η^2	Signif.
Room	$F(1.8, 18.3) = 52.71$, $p < 0.001$, η^2 =0.84	*
Talker config.	$F(1, 10) = 58.55$, $p < 0.001$, η^2 =0.85	*
Type of exposure	$p = 0.11$	
Room × Talker config.	$F(1.9, 19.0) = 37.77$, $p < 0.001$, η^2 =0.67	*
Room × Type of exposure	$p = 0.21$	
Talker config. × Type of exposure	$F(1, 10) = 23.56$, $p < 0.005$, η^2 =0.70	*
Room × Talker config. × Type of exposure	$p = 0.06$	

the omnibus ANOVA was not significant. The interaction between room and talker configuration was significant. The interaction between talker configuration and type of exposure was also significant. The main effects of room and talker configuration were significant.

Table 6.3: Results of two-way repeated-measures ANOVAs on the co-located and spatially-separated thresholds for the multi-talker speech identification task.

	Co-located	
Factor	F-, p-values, η^2	Signif.
Room	$F(1.9, 19.0) = 7.16$, $p < 0.005$, η^2 =0.42	*
Type of exposure	$p = 1.00$	
Room × type of exposure	$p = 0.10$	
	Spatially-separated	
Factor	F-, p-values, η^2	Signif.
Room	$F(1.7, 16.6) = 37.81$, $p < 0.0001, \eta^2 = 0.79$	*
Type of exposure	$F(1, 10) = 9.71, p < 0.05, \eta^2 = 0.49$	*
Room × type of exposure	$p = 0.12$	

The co-located and spatially separated threshold data were further analyzed with separate two-way repeated-measures ANOVAs with room and type of exposure as factors (4×2). The ANOVA results are summarized in Table 6.3.

There was a small but significant main effect of the room on the co-located threshold. Pairwise multiple comparisons indicated significant differences in the following pairs: R0 and R1 ($p = 0.005$) and R0 and R3 ($p = 0.001$). Thus, an increase in threshold occurred due to the addition of reverberation. However, the differences in the lateral spread of the reverberation (*cf.* between R1, R2, and R3) did not significantly affect the co-located thresholds. Furthermore, the main effect of exposure type and the interaction between room and exposure type were not statistically significant in the co-located condition.

Thresholds in the spatially separated condition increased in the order R0, R1, R2, R3. This was expected due to the corresponding increase in the lateral spread of reflections. Note that the benefit of talker separation in R1 was significantly lower than that in R0, even though other listeners' abilities in localizing a single source were similar in R0 and R1 (Experiment I, Chapter 4). The decrease may be attributed to increases in both energetic and informational masking. Furthermore, the increase in the spectrotemporal overlap of the competing sentences might further degrade other non-spatial segregation cues, making stream segregation more difficult and increasing informational masking.

ANOVA results for the thresholds in the spatially separated configuration indicated a significant main effect of room and a small but statistically significant main effect of the type of exposure. The interaction between room and type of exposure was not statistically significant. The average thresholds in the fixed-exposure condition were lower than the thresholds in the mixed-exposure condition except in R2. The differences in threshold between the mixed and fixed exposure conditions were on average 0.5 dB across all rooms and participants (0.9 dB, 0.9 dB, -0.2 and 0.3 dB in R0, R1, R2, and R3, respectively). This result suggests that listeners could gain more benefit from the location cues when the salience of such cues was consistent across trials.

The amount of spatial release from masking (SRM), defined as the threshold in the spatially-separated condition subtracted from the threshold in the co-located condition, is plotted in Fig. 6.4. A positive SRM value (i.e., data points above the dotted line in the figure) indicates an advantage of talker separation in this task. A large individual variability in SRM can be seen from the figure. Furthermore, SRM decreased with an increase in the lateral spread of the reverberation. However, it can be seen that most listeners obtained the benefit of talker separation in all rooms and types of exposure.

Additionally, the types of erroneous keywords reported (i.e., colour and number tokens that did not match those spoken by the target talker) were analyzed and are

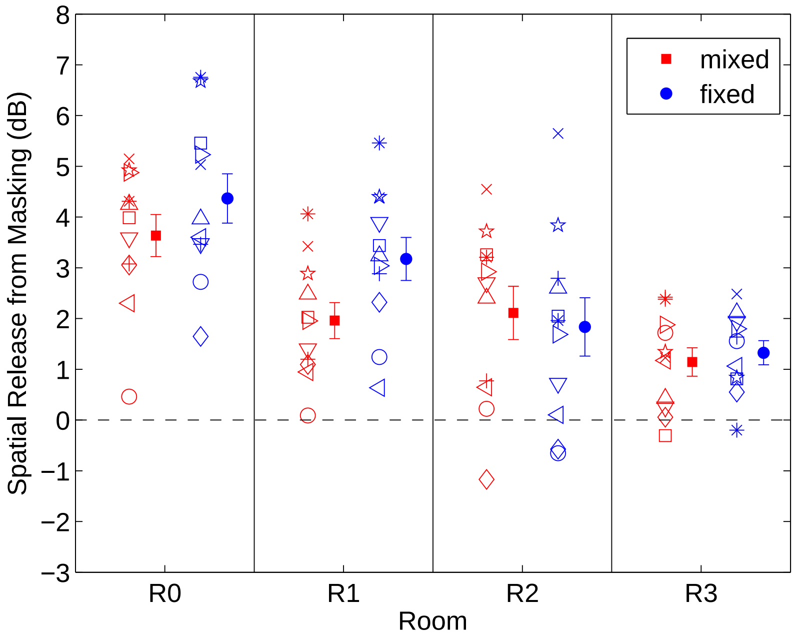


Figure 6.4: Spatial release from masking, i.e., the difference between the thresholds in the co-located and spatially separated conditions. Each marker type corresponds to an individual. Markers on the left (in red) and on the right (in blue) indicate data in the mixed- and fixed-exposure conditions, respectively. Filled markers are average data across listeners. Error bars indicate ±1 standard error of the mean.

shown in Fig. 6.5. Erroneous keywords were classified into one of the following types: those spoken by the left competing talker (blue bars in the figure), those spoken by the right competing talker (right bars), or those that were spoken by none of the talkers (green bars). Each keyword (colour or number) error was counted separately. For example, if a listener reported both the colour and number spoken by the left talker, 2 counts were added to the left-masker total, whereas if the reported colour matched with that spoken by the left talker and the reported number matched with that spoken by the right talker, 1 count was added to both the left-masker and right-masker totals. One count was added to the non-present totals for each reported colour or number that did not match any of those spoken by the target, left, or right talkers. Data from the 6 T/M conditions were summed for the analysis; hence, there were 6 T/Ms $\times$ 30 trials/$T/M \times$ 2 keywords/trial = 360 keywords in total for each condition. The data across listeners and types of exposure were averaged.

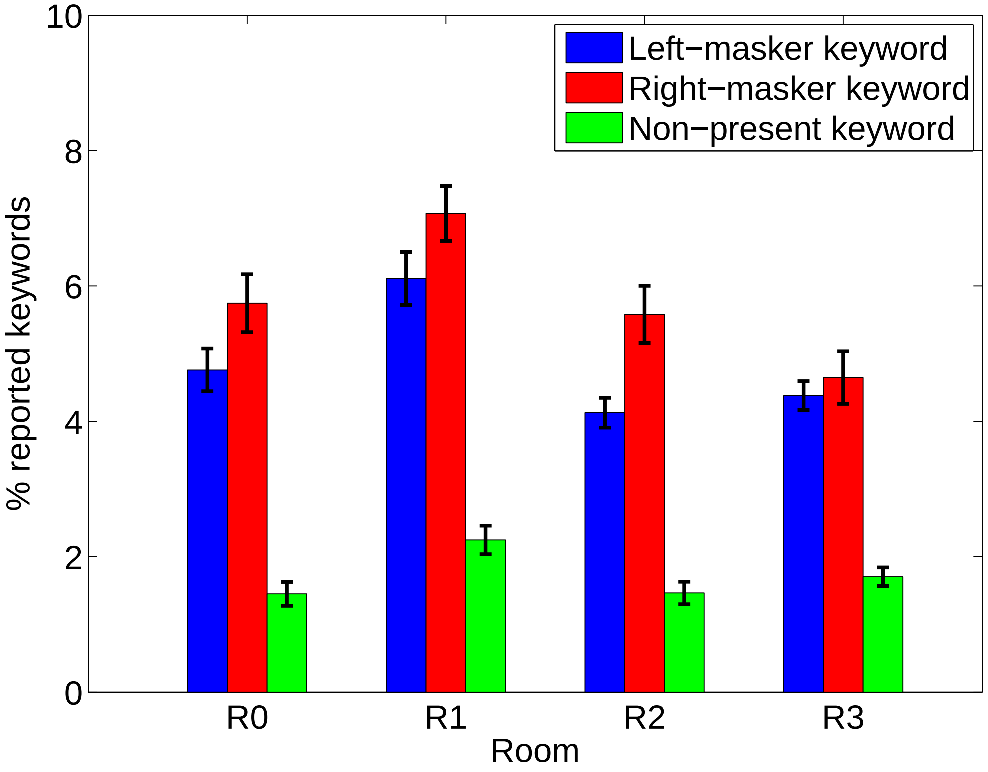


Figure 6.5: Percent of erroneous keywords (colour and number tokens) reported that corresponded to the keywords spoken by the left masker (-22.5°, blue bars), right masker (+22.5°, red bars), or neither of the maskers (green bars). Data were averaged across types of exposure and participants ($n = 11$).

From Fig. 6.5, it can be seen that there were many fewer non-present keyword errors than left- and right-masker errors. That is, listeners rarely reported keywords that were not present in a trial. Additionally, a higher tendency of reporting a colour or a number spoken by the right competing talker can be seen from the figure. The

Table 6.4: Two-way repeated-measures ANOVA on keyword errors.

Factor	F-, p-values, η^2	Signif.
Room	$F(2.5, 25.5) = 30.09$, $p < 0.001$, η^2 =0.75	*
Masker side	$F(1, 10) = 6.78$, $p < 0.05$, η^2 =0.40	*
Room × masker side	$p = 0.18$	

left- and right-masker error data (omitting non-present errors) were submitted to a two-way repeated measures ANOVA with room and masker side as factors (Table 6.4). The results indicated a significant main effect of masker side; more right-masker keywords than left-masker keywords were reported by the participants. Recall that an asymmetry favouring the right hemifield was also observed in Experiment I (horizontal sound localization; Chapter 4). In Experiment I, front/back reversals occurred more often when the target originated from the left hemifield.

6.2.3 Discussion

Results from the speech identification experiment indicated that providing consistent reverberation context yielded a slight advantage for correctly identifying the target keywords. However, this advantage was only observed in the spatially separated talker configuration, indicating that the consistency of reverberation further enhanced spatial release from masking. The lack of significant effect of exposure in the co-located condition suggests that listeners were not able to improve their ability to take advantage of other cues that could be used to direct selective attention to the target sentence (for example, differences in talkers' voices or speech levels).

The finding is also consistent with the theory that selective attention builds up over time (Kahneman, 1973) and that the continuity in the features of an attended object facilitates the buildup of selective attention (Best et al., 2008). As mentioned in the Introduction of this chapter, Best et al. (2008) showed that continuity in the target location facilitated the buildup of selective attention, whereas changes in the target location incur cost in selective attention, even when advance visual cues are presented. In our experiment, the target location was fixed at 0°, and the listeners were aware of its location and of the spatial configuration of the competing talkers. However, the variability of the reverberation (in this case, the variability in interaural coherence) in the mixed exposure condition might have forced listeners to re-evaluate the auditory scene on each trial, and thus, prevent the build up of spatial-based selective attention.

The results of this experiment suggests that listeners can take advantage of consistency in reverberation context when there exists spatial separation among the talkers. Additionally, prior knowledge of the target location and its continuity across trials may have contributed to the exposure effect. The next experiment aimed to explore whether consistency of interaural coherence can also facilitate listeners' ability to *locate the position* of a target talker within a multi-talker configuration.

6.3 Experiment IIIB: The effect of consistency in reverberation on target location identification

6.3.1 Method

Participants

Ten young, normally hearing listeners who were fluent in English participated in this experiment (2 males, 8 females; 20-30 years of age, mean age = 25). Five participants (S03, S08, S13, S14, and S18) participated in Experiment IIIA. Two participants took part in Experiment I (S03, S24), and 1 participant (S18) took part in Experiment II.

Testing was typically performed in 2×1.5-h sessions. Testing progressed according to individual pace and schedule. Short breaks were provided whenever needed.

Stimuli and procedure

In this experiment, participants were simultaneously presented with CRM sentences from three same-sex talkers. The talkers were positioned at -22.5°, 0°, and +22.5°. One of the sentences was designated as the target, and had the callsign “Baron”. The other two sentences had callsigns other than “Baron”, and all colours and numbers differed from one another. The target sentence was equally likely to be presented from any of the three possible positions.

Participants were asked to identify the location of the target speech by pressing one of three buttons displayed on a touchscreen monitor; they were labeled “Left”, “Centre”, and “Right”, indicating locations at -22.5°, 0°, and +22.5°, respectively. A response had to be provided in each trial, and the participants were asked to guess if needed. No feedback on the correct response was provided.

There were 8 listening conditions in total, corresponding to the combination of 2 types of exposure (mixed and fixed) and 4 rooms (R0, R1, R2, and R3). Additionally, the target sentence was presented at 2 T/Ms of 6 and 4 dB in separate blocks of trials. These T/M values were chosen based on a pilot experiment to give performance above chance level and below ceiling performance in each of the rooms and across room conditions. There were 90 trials in total for each listening condition and T/M combination, corresponding to 30 trials from each of the three possible locations. For each T/M, there were 8 blocks of 45 trials in both the mixed- and fixed-exposure conditions.

Five of the participants completed the mixed-exposure condition followed by the fixed-exposure condition. The other five participants performed the experiment in the reverse order. The order of the rooms in the fixed-exposure condition was randomized for each participant. Furthermore, the 6-dB T/M blocks were always presented before the 4-dB T/M blocks. Typically, the mixed and fixed blocks were presented in different sessions. Prior to starting the experiment, each listener was presented with 45 R0-trials with a T/M of 8 dB as practice.

6.3.2 Results

Results from the 6-dB and 4-dB trials were combined, and the percentages of correct responses were calculated. The percentage of correct response averaged across individuals and target locations is plotted in Fig. 6.6. The effect of the room can clearly be seen from this figure. As expected, performance decreased when the reverberation was added and as the lateral spread of reflections was increased.

The percent-correct data, partitioned according to room and target location are plotted in Fig. 6.7. These data were were transformed into rationalized arcsine units (Studebaker, 1985, RAU) to make them more suitable for ANOVA, given the near-ceiling effects obtained in some conditions. The RAU data were then submitted to a three-way (4×2×3) repeated-measures ANOVA (room, type of exposure and location as factors, respectively). The three-way interaction in this omnibus ANOVA was significant [$F(3.78, 34.04) = 3.34, p < 0.05, \eta^2 = 0.271$].

Given the three-way interaction in the omnibus ANOVA, a separate two-way repeated-measures ANOVA was applied to the data in each room for ease of interpretation (αs were adjusted with Šidák correction factor). The ANOVA results are

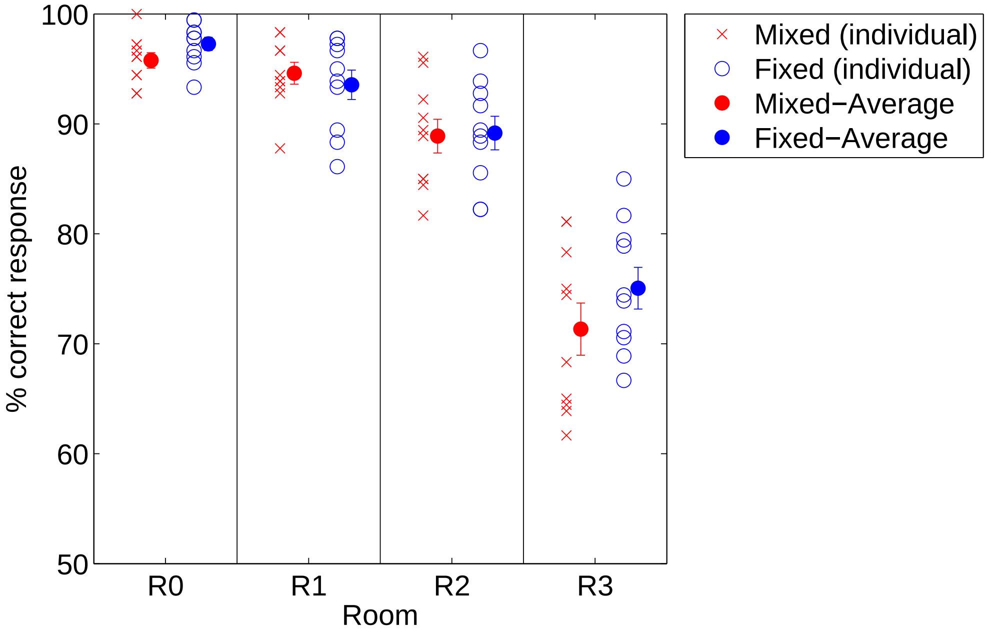
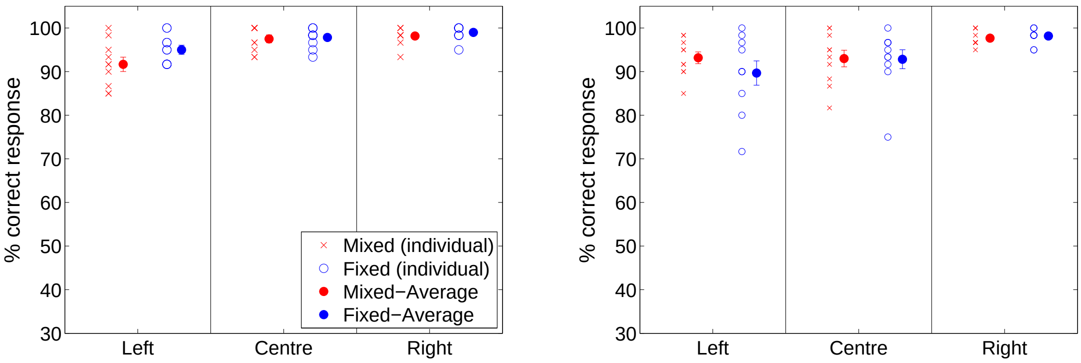


Figure 6.6: Individual and average percentages of correct responses for the target location identification task in R0, R1, R2, and R3 ($n = 10$). Error bars indicate ±1 standard error of the mean.

shown in Table 6.5. In all rooms, the main effect of exposure was not significant. The main effect of target location was significant in all rooms. However, the main effect of exposure was not significant in any of the rooms. Only in R3 was the two-way interaction (location × type of exposure) significant. Pairwise t-tests applied to test the significance of the type of exposure at each target location in R3 indicated that the mixed-fixed difference was significant only for the centre target $[t(9) = 4.00, p < 0.005]$ (centre panel of Fig. 6.7d).

As can be seen in Fig. 6.7a-d, in general, listeners were more successful in locating the target sentence when it was positioned at +22.5°. This asymmetry (with respect to -22.5°) might be caused by a possible perceptual asymmetry or by acoustical asymmetries in the simulated reverberation.

(a) R0 **(b)** R1



(c) R2 **(d)** R3

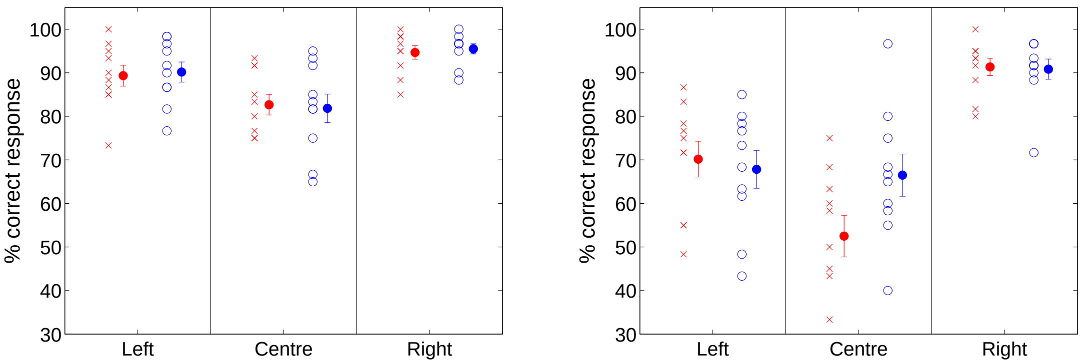


Figure 6.7: Performance in the the target location identification task (% correct response) for each target location in: (a) R0, (b) R1, (c) R2, and (d) R3. *Left*, *Centre*, and *Right* correspond to -22.5°, 0°, and +22.5°, respectively. The legends for all figures are identical to those shown in (a). Error bars are ±1 standard error of the mean.

To evaluate possible acoustic causes of left-right asymmetry in this task, we carried out additional experiments. Five additional participants (including S24, who participated in Experiment IIIA and the author) were tested in the R0, fixed-exposure condition facing the 180°-loudspeaker (see Fig. 3.1 in Chapter 3) at T/Ms of 6, 4, 2, 0, and -2 dB. This condition tests the listening environment (i.e., the anechoic chamber) and the loudspeaker setup itself. To evaluate acoustical asymmetries, four additional

Table 6.5: Two-way repeated-measures ANOVA tables for the RAU data on the target location identification task in R0, R1, R2, and R3.

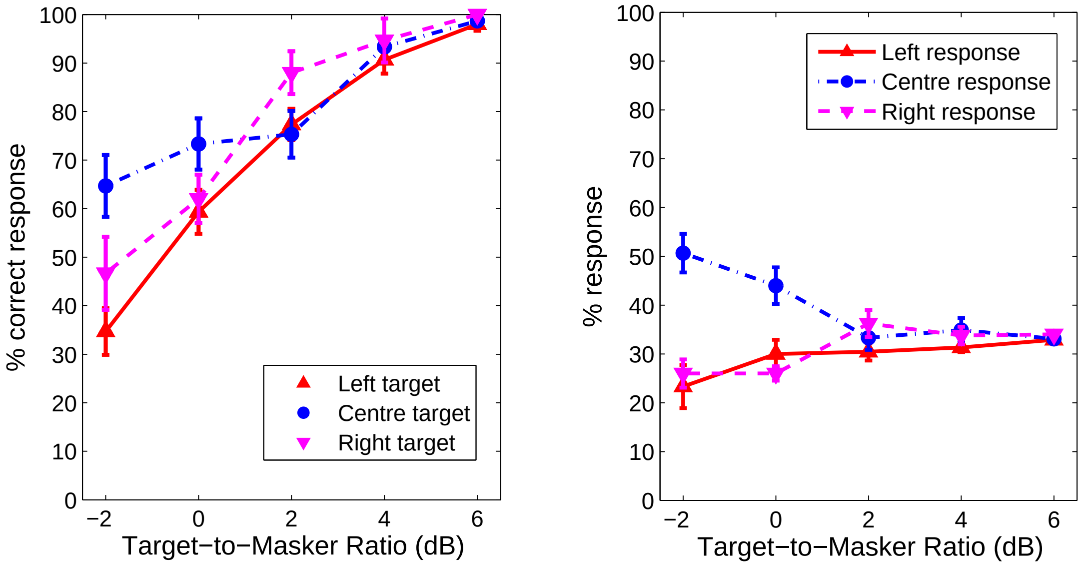
	R0	
Factor	F-, p-values, η^2	Signif.
Location	$F(1.5, 13.1) = 14.36$, $p < 0.005$, η^2 =0.615	*
Type of exposure	$p = 0.11$	
Location × type of exposure	$p = 0.38$	
	R1	
Factor	F-, p-values, η^2	Signif.
Location	$F(1.7, 15.7) = 7.99$, $p < 0.01, \eta^2 = 0.47$	*
Type of exposure	$p = 0.502$	
Location × type of exposure	$p = 0.482$	
	R2	
Factor	F-, p-values, η^2	Signif.
Location	$F(1.8, 16.4) = 12.83$, $p < 0.005, \eta^2 = 0.59$	*
Type of exposure	$p = 0.810$	
Location × type of exposure	$p = 0.865$	
	R3	
Factor	F-, p-values, η^2	Signif.
Location	$F(1.7, 15.6) = 21.98$, $p < 0.001, \eta^2 = 0.79$	*
Type of exposure	$p = 0.130$	
Location × type of exposure	$F(1.5, 13.7) = 8.74, p < 0.01, \eta^2 = 0.49$	*

participants (including S18, who participated in Experiment IIIA) performed the a condition similar to the R3, fixed-exposure condition at 6 dB, but with the sidedness of the reflections in the stimuli reversed. That is, the +45°- and +135°-reflections were assigned to the -45° and -135° loudspeakers, respectively, and vice versa. All 9 participants in these additional experiments, as well as the 10 participants who took part in the original experiment, reported that they were right-handed. The results are shown in Fig. 6.8. Additionally, contingency tables showing the target and response patterns, averaged across listener, for the R0, 180°-facing experiment at each T/M are shown in Table 6.6.

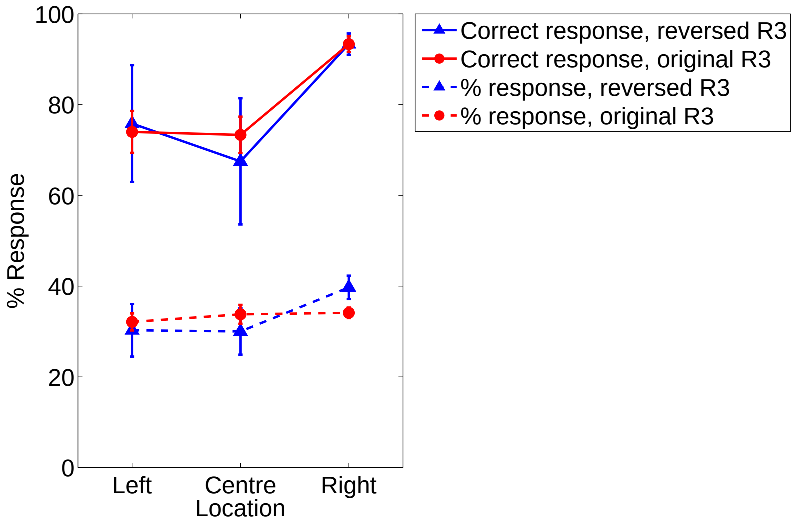
The patterns observed in these data were similar to those demonstrated in the original location identification experiment. In Fig. 6.8a (R0, facing 180°), the correct response rates were higher when the target was located at the right (+22.5°) at 6, 4, and 2 dB. Correct response rates were highest for the centre-located target at T/Ms lower than 2 dB. However, this was likely due to listeners' tendency to choose a "Centre" response when they could not detect the target callsign. This is indicated by an increase in the rate of "Centre" responses as the T/M was decreased (Fig. 6.8a, right panel). For the left-right reversed R3 condition at 6 dB T/M (Fig. 6.8b), the pattern of response was similar to that obtained in the original experiment; target sentences at +22.5° were more likely to be correctly located than those at the other locations despite the reversed laterality of the reflections. t-tests on the percent-correct data comparing the performance in the original R3 and left-right reversed R3 conditions indicated that none of the differences was statistically significant [Left target: $t(12) = 1.52, p = 0.154$; Centre target: $t(12) = 1.49, p = 0.162$; Right target: $t(12) = 1.64, p = 0.128$].

6.3.3 Discussion

In contrast to the finding in the speech identification experiment (Experiment IIIA), in Experiment IIIB listeners performed similarly in both the mixed and fixed exposure conditions at the T/Ms tested (6 and 4 dB). That is, their ability to locate the target sentence was not improved by the consistency of the reverberation. This finding also contradicted the exposure effect found in the single-source sound localization task (Experiment I). Perhaps the adaptation mechanisms involved in the single-source localization task failed to occur in the presence of masking energy or under higher levels of task demand.



(a) Target location identification performance in fixed, R0 condition with the listeners facing the 180° loudspeaker. Left: percentage of correct responses (out of 30 trials in each position and T/M; $n = 5$). Right: the percentage a particular location was chosen as a response (out of 90 trials).



(b) Percentages of correct target-location responses in the fixed, R3 condition with the left/right sidedness of the reflections reversed (blue markers, solid line) at T/M of 6 dB T/M (n=4). Plotted in the red markers (solid line) are the data from the original experiment at 6 dB T/M ($n = 10$). Data connected with the dashed lines are the percentages of left, centre, and right responses (from left to right, respectively).

Figure 6.8: Performance in the additional conditions in the location identification task: (a) in the fixed-R0 condition with a different set of loudspeakers, (b) in the fixed-R3 condition with the sidedness of the reflections reversed.

Table 6.6: Target/response contingency tables for the location identification task in the R0, 180°-facing condition at T/Ms of 6, 4, 2, 0, and -2 dB. Each entry in the table is the total number of Left/Centre/Right responses (out of 30 trials) given Left/Centre/Right targets. Data were averaged across listeners ($n = 5$).

	T/M = 6 dB			T/M = 4 dB			T/M = 2 dB		
Target	Response			Response			Response		
	Left	Centre	Right	Left	Centre	Right	Left	Centre	Right
Left	29.4	0.2	0.4	27.2	2.2	0.6	23.2	4.8	2
Centre	0.2	29.6	0.2	0.6	28	1.4	3.2	22.6	4.2
Right	0	0	30	0.4	1.2	28.4	1	2.6	26.4

	T/M = 0 dB			T/M = -2 dB		
Target	Response			Response		
	Left	Centre	Right	Left	Centre	Right
Left	17.8	9.2	3	10.4	14.4	5.2
Centre	6.2	22	1.8	6.4	19.4	4.2
Right	3	8.4	18.6	4.2	11.8	14

There were a number of limitations to the design of the current study. First, the test was conducted only at two T/Ms (6 and 4 dB), that likely provided little or no informational masking. The pattern of response may differ under lower target detectability. Thus, the results obtained in the current study should not be generalized to other T/M-conditions. Second, given the 3-AFC design, listeners were required to choose a response even when the target was not audible to them. In the current study, most listeners tended to choose “Centre” as a preferred response as the target callsign became harder to detect. Thus, the results became less reliable as the T/M was decreased. The design could have been improved by allowing the listeners to indicate when the target was not detected.

The results from this experiment also suggest a right-hemifield advantage in locating the target sentence; listeners were more successful in correctly identifying the target location when it was positioned at +22.5° (Fig. 6.7). Additional experiments were performed to evaluate whether asymmetries in the experimental setup or the simulated reverberation might have contributed to this asymmetry. The results demonstrated patterns of responses similar to those obtained in the original experiment (Fig. 6.8), indicating that neither the acoustics of the testing room nor of the simulated reverberation were responsible for the asymmetry in the results. Right-hemifield advantages have also been reported for target detection and speech identification tasks using the CRM materials and talker configurations (symmetrically positioned competing talkers) similar to that used in the current study (Bolia, Nelson, & Morley, 2001; Kidd, Arbogast, et al., 2005).

6.4 Summary of results and general discussion

The results of Experiments IIIA and IIIB can be summarized as follows:

- Reverberation significantly reduced the benefit of talker separation in the multi-talker speech identification task (Experiment IIIA). However, significant spatial release from masking could still be observed in all of the reverberation conditions in our study.

- Speech identification performance was better in the consistent reverberation condition only when the talkers were spatially separated. Reverberation context did not have a significant effect on performance when the talkers were co-located. This suggests that consistency in reverberation enhances spatial release from masking.

- No significant effect of exposure was found in the location identification experiment (Experiment IIIB). However, the tasks were conducted only at high

T/Ms (6 and 4 dB) so that little informational masking was present. Further experiments must be conducted to evaluate the effect of exposure under lower levels of target detectability.

- A right-hemifield advantage was observed in the location identification task. In the speech identification task, a higher number of erroneous keywords reported corresponded to those spoken by the right masker than those spoken by the left masker. Results from additional experiments suggested that the response asymmetry in the location identification task was likely not caused by stimulus asymmetries.

The results of Experiment IIIB suggest that consistency in reverberation did not help listeners' ability in locating the target talker in the multi-talker situation. In Experiment IIIA, the target speech was always presented from the front, and the listeners were aware of the spatial configuration of the talkers. A possible explanation for the exposure effect obtained in Experiment IIIA is that the fixed target location facilitates listeners' ability to maintain spatial selective attention. Consistent reverberation may facilitate the build up of spatial selective attention, whereas the inconsistent reverberation in the mixed-exposure condition may cause listeners to re-evaluate the auditory scene on each trial, preventing the buildup of spatial selective attention towards the target location.

Another alternative explanation for the exposure effect obtained in Experiment IIIA is that a consistent reverberation context may decrease the perceptual variability of the target talker location in Experiment IIIA, as in Experiment I. However, the measure used in the target location identification would only have been sensitive to very large changes in localization precision.

The response asymmetry found in both Experiments IIIA and IIIB provided additional supporting evidence to the findings of other studies that also observed a similar pattern of asymmetry in multi-talker listening situations (Bolia et al., 2001; Kidd, Arbogast, et al., 2005). This warrants further investigation, and future research could be conducted to assess the applicability of the results. Better understanding of this phenomenon could be beneficial for improving auditory display systems, e.g., presenting important messages from the right hemifield in appropriate situations, or for improving listening strategies in multi-talker listening situations.

Chapter 7

Conclusions

Overall, the results of Experiments I, II, and III indicated that the effect of reverberant context on spatial hearing performance was dependent upon the task. The benefit was observed in a single-source localization task (Experiment I) and in a speech perception task with a multi-talker configuration in which the talkers were spatially separated (Experiment IIIA). The effect of reverberation context, however, was not statistically significant in a speech-in-noise task (Experiment II) and in a target location identification task in a multi-talker configuration (Experiment IIIB).

The following sections present a summary of the main findings of Experiments I, II, and III, as well as their limitations (Sec. 7.1), a summary of the contributions from this thesis (Sec. 7.3), and suggestions for future research (Sec. 7.4).

7.1 Summary of findings and limitations

7.1.1 Experiment I

In Experiment I, listeners' judgements on the azimuth positions of reverberant targets (two-word speech tokens) were measured in the mixed and fixed reverberant contexts in four simulated rooms (R0, R1, R2, and R3). The main findings from this experiment were:

1. The variability in listeners' azimuth judgements of R1-, R2-, and R3-targets was smaller in the fixed reverberation context than in the mixed reverberation context. The difference was relatively small; almost all listeners, however, achieved lower response variability in the fixed exposure condition. This result provides support to the hypothesis that sound localization depends not only on the cues present in isolated trials, but also on the pattern of cues in the stimulus context.
2. Steady-state ILD cues had little effect on listeners' azimuth responses. ILD cues in R3 were clearly biased towards the midline. However, mean azimuth judgements of R3-stimuli were not significantly different from those of R0-stimuli (free-field stimuli). This result is in accordance with the known dominance of ITD and onset cues in horizontal sound localization.

Evidence of dependence of localization judgements on stimulus context necessitates a model of sound localization that takes into account the statistics of past trials. Such a model may perhaps employ an adaptive cue weighting strategy that optimizes cue weights across frequency and time based on information in past trials. An example of such a model is the adaptive cue selection strategy conceptually proposed by Faller and Merimaa (2004) to model the direction-dependent buildup of the precedence effect.

A limitation of Experiment I was that its design could not yield further insight into the potential mechanisms underlying the effect of reverberation context. Perhaps spatial cues across frequency and time were weighted differently according to the reverberation context. Future sound localization or lateralization experiments employing carefully designed stimuli that allow for quantification of relative cue-weights across frequency and/or time should be conducted to evaluate this cue re-weighting hypothesis.

7.1.2 Experiment II

Experiment II aimed to examine whether a mechanism related to compensation for degraded interaural coherence played any role in the enhancement of speech intelligibility found in several previous studies (Brandewie & Zahorik, 2010; Srinivasan & Zahorik, 2013; Zahorik & Brandewie, 2011). An additional aim of Experiment II was to examine the contributions of the presence of a preceding sentence carrier and across-trial consistency to the exposure effect obtained in those studies. The main findings from Experiment II were:

1. Speech intelligibility performance decreased with an increase in the lateral spread of reflections. This could be expected from the degradation in the acoustic cues (i.e., the loss of better-ear advantage and binaural squelch).

2. Unlike the finding of Experiment I which suggests that listeners gain benefit from across-trial exposure to stimuli with consistent reverberation, no significant effect of reverberation context was found in this experiment. Similar speech intelligibility performance ($\mathrm{SNR}_{71\%}$) was obtained in both the mixed and fixed reverberation contexts (*cf.* Mixed-Short and Fixed-Short conditions, respectively).

3. The effect of the sentence carrier (*cf.* difference in intelligibility between the Fixed-Short and Fixed-Full conditions) was not statistically different across

rooms (R0, R1, R2, and R3). The amount of improvement due to the sentence carrier was on average larger in R1, R2, and R3 than in R0, similar to the pattern observed in other related studies (Brandewie & Zahorik, 2010; Zahorik & Brandewie, 2011). However, the sizes of the effect in R1, R2, and R3 were smaller than those obtained in similar studies performed by other researchers. Furthermore, they were not statistically different across all rooms.

The lack of a significant effect of reverberation context and the lack of a significant interaction between sentence-carrier and room in our study suggest that listeners did not gain a significant advantage from having consistent reverberation context when only the interaural coherence was varied. Perhaps similar to related studies (Watkins et al., 2011; Srinivasan & Zahorik, 2014), the effects of reverberation context observed in the studies of Brandewie and Zahorik (2010) and Zahorik and Brandewie (2011) were due to the context's amplitude envelope rather than to its interaural coherence.

Averaged across listeners, the advantage of sentence carrier for R0-stimuli in our study was 0.54 dB, roughly comparable to the 0.84 dB obtained by Brandewie and Zahorik (2010). However, the sentence-carrier advantages for the reverberant stimuli obtained in our study were smaller than the 2.7-dB advantage obtained in the latter study (the sentence-carrier advantages obtained in our study were 0.98, 1.53, and 1.30 dB in R1, R2, and R3, respectively). Additionally, there was a considerable amount of inter-individual variability in our data (Fig. 5.4). The combination of small effect sizes and the amount of individual variability would have made it difficult to detect differences in effect size across room conditions. Given that the advantage of exposure seemed to increase with increasing reverberation time for $0.3s < T_{60} < 1.2s$, future experiments would benefit from using rooms with a reverberation time longer than that used in the current experiment ($T_{60} \approx 0.47$ s).

7.1.3 Experiment III

Experiment III examined the effect of reverberation context on listeners' ability to identify target speech (Experiment IIIA) and to identify the location of a target talker (Experiment IIIB) in listening situations involving three simultaneous same-sex talkers. In Experiment IIIA, speech intelligibility was tested under co-located (all talkers at 0°) and spatially separated (talkers at -22.5°, 0°, and +22.5°; the target was always at 0°) talker configurations. In Experiment IIIB, listeners were asked to identify the location of the talker uttering the target callsign "Baron" amongst three talkers positioned at -22.5°, 0°, and +22.5°. The main findings of these experiments were:

1. For the co-located talker configuration, speech intelligibility performance was similar in both the mixed and fixed reverberation contexts. For the spatially separated talker configuration, the effect of reverberation context was statistically significant; speech intelligibility performance was better in the fixed exposure condition than in the mixed exposure condition. This result suggests that a consistent coherence context improves spatial release from masking.

2. Performance in the location identification task was similar in both the mixed and fixed exposure conditions. Thus, if listeners did not have a priori information about the target location in Experiment IIIA, it would have been less likely to observe a context effect as their ability to locate and direct their attention by the target's location was not improved by the consistent reverberation context. A possible explanation for the effect obtained in Experiment IIIA was that the

consistency in reverberation facilitated the buildup of selective spatial attention towards the target location, while changes in the reverberation caused listeners to re-evaluate the auditory scene on each trial. Alternatively, a consistent reverberation context might reduce the perceptual variability of target talker location. However, such an improvement could not have been detected with the measure used in the location identification task (Experiment IIIB).

3. Asymmetry in responses favoring the right-hemifield targets was found in both Experiments IIIA and IIIB. Results from follow-up experiments indicated that the response asymmetry was likely not caused by stimulus asymmetries. Other studies have also reported a similar pattern of asymmetry using the CRM material in talker configuration similar to that used in our study (Bolia et al., 2001; Kidd, Mason, & Brughera, 2005).

A limitation of Experiment IIIB was that the task was conducted only at relatively high T/Ms of 6 and 4 dB. Further experiments could be performed to evaluate the effect of exposure under lower target audibility. Furthermore, the 3-AFC procedure in this experiment caused listeners to increase their “Centre” response as the target audibility was lowered. Future experiments would benefit from having the listeners indicate the events of which the target could not be detected.

The finding related to asymmetry favoring stimuli presented from the right hemifield warrants further investigation, as would further research to assess the applicability of this finding in practical applications. For example, an auditory display system could present important messages from the right side.

7.1.4 Limitations

The experiments in this thesis were conducted in a simulated reverberation environment which in several ways did not reflect the typical characteristics of reverberation found in real rooms. First, in real rooms, reflection density builds over time, while for the simulated reverberation used in our studies, the impulses in the late part of the reverberation were equally spaced by 4 ms. The train of random-polarity, decaying amplitude, 4-ms spaced impulses, was, however, intended to model the amplitude envelope and noise-like characteristics of a reverberation tail, not individual reflections. Replacing discrete reflections with a noise-like random process beginning at approximately 50 ms has also been used in several studies to simulate the late part of the reverberation (e.g., Zahorik, 2009; Bradley et al., 2003). Second, interaural coherence decreases with time in real rooms, in part due to the build up of the reverberant energy (which would have occurred similarly in our simulated rooms), but also due to the spatial distribution of the reflections becoming increasingly diffuse with time; in our simulated reverberation, the reflections in the early and late parts of the reverberation had the same spatial distribution. Third, real-room reverberation typically possesses low-pass characteristics; the energy of high-frequency components is absorbed more readily than the energy of low-frequency components. The simulated reverberation employed in our experiments applied equal gain to all frequency components in the reflections. It is possible that the exposure effects obtained in other studies depend on the characteristics of reverberation found in real rooms, thus contributing to the small effect sizes found in our studies. However, the simulated reverberation captured a few essential characteristics of real-room reverberation, namely, a loss of modulation depth in the signal's amplitude envelope attributed to the reverberation time and a decrease in interaural coherence. It is difficult to vary these parameters independently in real rooms. Utilizing a simulated reverberation rather than real-room reverberation allowed us to exert more control on these parameters and vary them independently (i.e., to vary interaural coherence while maintaining reverberation time).

7.2 General Discussion

The studies presented in this thesis have investigated the effect of reverberation context on a number of spatial hearing tasks, in particular whether consistency in interaural coherence could be beneficial for listeners' performance. The results from these studies indicated that consistent coherence context could be advantageous in the single-source sound localization task (Experiment I) and spatial release from masking in a multi-talker configuration (Experiment IIIA). The effect of coherence context was not significant in the other tasks tested, namely, speech perception in a spatially separated noise (Experiment II), and target location identification in multi-talker configurations (Experiment IIIB).

The finding of a significant context effect in the single-source sound localization tasks suggests that listeners could account for the pattern of stimuli in the context and that less variability in the context allowed them to use a more optimal strategy. Perhaps similar to the buildup of the precedence effect, listeners could increase the perceptual weights of reliable spatial cues as evidence about the stimulus pattern accumulated. Further studies, however, need to be conducted to evaluate this hypothesis.

The benefit of consistent coherence context in Experiment I, however, was small and was not observed in Experiment II and IIIB. A possible explanation for the lack of significant effect in these tasks is that the masker energy further degraded the interaural coherence, reducing the differences in interaural coherence across the reverberant conditions.

In other studies similar to Experiment II (Brandewie & Zahorik, 2010; Zahorik & Brandewie, 2011) which also examined the effect of reverberation context employing reverberation with various reverberation times (ranging from 0.3 s to 3 s) and varying coherence, a significant effect of reverberation context was found. The result of Experiment II, however, suggests that the effect obtained in these studies was likely not due to consistency in coherence. Perhaps the effect observed in other studies could be attributed mainly to consistency in the context's amplitude envelope, as the results of several related studies suggest (Watkins et al., 2011; Srinivasan & Zahorik, 2014).

A small but significant effect of reverberation context was found in the multi-talker speech perception experiment (Experiment IIIA), but only when the talkers were spatially separated. Experiment IIIB, however, indicated that listeners' ability to identify the location of the target talker in a similar multi-talker configuration was not improved with consistent reverberation context, suggesting that the context effect in Experiment IIIA was not due to improved localizability of the target. A possible explanation of the context effect observed in Experiment IIIA is that consistency of reverberation facilitated the buildup of selective spatial attention towards the fixed target location.

The studies presented in this thesis have examined the effect of reverberation context over a relatively short duration (each of the room-specific, fixed exposure condition was typically completed in shorter than the duration a test session); nevertheless, the results of some of these studies indicate that stimulus pattern within a context can influence listeners' responses. For sound localization, it has been shown that listeners' accuracy in localizing sounds in a room continues to improve over the duration of several days of testing (3-5 days) even in the absence of feedback (Shinn-Cunningham, 2000). Further studies could be conducted to examine the effects of reverberation context over a longer duration of testing.

7.3 Contributions

The studies presented in this thesis provided additional evidence of the effect of reverberation context on spatial hearing performance. In particular, the role of interaural coherence in the context was examined. The findings of these studies provide additional insight on the potential mechanisms that may contribute to the effect of reverberation context found by other researchers. That is, consistency in interaural coherence likely did not contribute to the reverberation context effect found in other studies.

We have also demonstrated the effect of reverberation context on spatial release from masking in multi-talker listening configurations. To our knowledge, no previous study has examined this issue. Additionally, our results suggest that while the ability to localize a speech target in multi-talker situations was not improved by consistent reverberation, consistent reverberant context could improve listeners' ability to selectively attend to a target speech based on its fixed location.

These studies were also the first to look at the effects of reverberation context in sound fields, in contrast to other studies which are conducted under headphones.

7.4 Future work

Further investigations of the mechanisms contributing to the exposure effect found in the single-source sound localization task could be performed. The hypothesis discussed earlier that perceptual cue weighting strategies are dependent on the reverberation context could be evaluated.

Furthermore, the effect of reverberation context examined in the present studies was measured over a relatively short time period (all fixed exposure, R_i stimuli were presented within a test session). Longer term exposure to room reverberation (over 3-5 days of listening experience in a room) has been shown to improve sound localization performance in the particular room (Shinn-Cunningham, 2000). Future research may explore the effect of reverberation context on performance in spatial hearing tasks over several days or weeks of testing sessions; for example, whether changes in performance can be observed under inconsistent/consistent reverberation context, or whether the rate of change in performance depends on reverberation context. Additionally, it would be of interest to examine the effectiveness of training with a consistent or inconsistent context, as well as the generalizability of training in a context to other contexts. This knowledge would be beneficial for developing auditory training programs to improve spatial hearing performance in reverberation.

References

Allen, J. B., & Berkley, D. A. (1979). Image method for efficiently simulating small-room acoustics. *J. Acoust. Soc. Am*, *65*(4), 943-950.

Amitay, S., Hawkey, D., & Moore, D. (2005). Auditory frequency discrimination learning is affected by stimulus variability. *Perception & Psychophysics*, *67*(4), 691-698.

ANSI. (1997). *Methods for the calculation of the speech intelligibility index* (No. S3.5).

Arweiler, I., & Buchholz, J. M. (2011). The influence of spectral characteristics of early reflections on speech intelligibility. *J. Acoust. Soc. Am*, *130*, 996-1005.

Aspeslagh, S., Clark, D. F., Akeroyd, M. A., & Brimijoin, W. O. (2014). Speech intelligibility can improve rapidly during exposure to a novel acoustic environment. *J. Acoust. Soc. Am*, *135*(4), 2227-2227. doi: http://dx.doi.org/10.1121/1.4877286

Best, V., Gallun, F. J., Carlile, S., & Shinn-Cunningham, B. G. (2007). Binaural interference and auditory grouping. *J. Acoust. Soc. Am*, *121*(2), 1070-1076. doi: http://dx.doi.org/10.1121/1.2407738

Best, V., Ozmeral, E. J., Kopčo, N., & Shinn-Cunningham, B. G. (2008). Object continuity enhances selective auditory attention. *Proceedings of the National*

Academy of Sciences, *105*(35), 13174-13178.

Blauert, J. (1997). *Spatial hearing - the psychophysics of human sound localization.* The MIT Press.

Blauert, J., & Col, J.-P. (1992). Auditory physiology and perception. In Y. Cazals, K. Horner, & L. Demany (Eds.), (p. 531-538). Oxford: Pergamon.

Bolia, R. S., Nelson, T., & Morley, R. M. (2001). Asymmetric performance in the cocktail party effect implications for the design of spatial audio displays. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *43*, 208.

Bolia, R. S., Nelson, W. T., Ericson, M. A., & Simpson, B. D. (2000). A speech corpus for multitalker communications research. *J. Acoust. Soc. Am*, *107*, 1065-1066.

Bradley, J. S., Sato, H., & Picard, M. (2003). On the importance of early reflections for speech in rooms. *J. Acoust. Soc. Am*, *113*, 3233-3244.

Bradley, J. S., & Soulodre, G. A. (1995). The influence of late arriving energy on spatial impression. *J. Acoust. Soc. Am*, *97*(4), 2263-2271.

Brandewie, E., & Zahorik, P. (2010). Prior listening in rooms improves speech intelligibility. *J. Acoust. Soc. Am*, *128 (1)*, 291-299.

Brandewie, E., & Zahorik, P. (2013). Time course of a perceptual enhancement effect for noise-masked speech in reverberant environments. *JASA Express Letters*, *134*(2), EL265-EL270.

Bregman, A. S. (1990). *Auditory scene analysis: the perceptual organization of sound.* The MIT Press.

Bronkhorst, A. W., & Plomp, R. (1988). The effect of headinduced interaural time and level differences on speech intelligibility in noise. *J. Acoust. Soc. Am*, *83*(4), 1508-1516.

Brown, A. D., & Stecker, G. C. (2013). The precedence effect: Fusion and lateralization measures for headphone stimuli lateralized by interaural time and level differences. *J. Acoust. Soc. Am*, *133 (5)*, 2883-2898.

Brungart, D. S., Simpson, B. D., Ericson, M. A., & Scott, K. R. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. *J. Acoust. Soc. Am*, *110 (5)*, 2527-2538.

Cherry, C. E. (1953). Some experiments on the recognition of speech, with one and with two ears. *J. Acoust. Soc. Am*, *25*, 975979.

Chiang, Y.-C., & Freyman, R. L. (1998). The influence of broadband noise on the precedence effect. *J. Acoust. Soc. Am*, *104*(5), 3039-3047.

Clifton, R. K., Freyman, R. L., & Meo, J. (2002). What the precedence effect tells us about room acoustics. *Perception and Psychophysics*, *64(2)*, 180-188.

Creelman, C. D. (1973). Simultaneous adaptive threshold estimation with frequency uncertainty. *J. Acoust. Soc. Am*, *54*(1), 316-316.

Culling, J. F., Hodder, K. I., & Toh, C. Y. (2003). Effects of reverberation on perceptual segregation of competing voices. *J. Acoust. Soc. Am*, *114*(5), 2871-2876.

Darwin, C. J., & Hukin, R. W. (1997). Perceptual segregation of a harmonic from a vowel by interaural time difference and frequency proximity. *J. Acoust. Soc. Am*, *102*(4), 2316-2324.

Darwin, C. J., & Hukin, R. W. (1998). Perceptual segregation of a harmonic from a vowel by interaural time difference in conjunction with mistuning and onset asynchrony. *J. Acoust. Soc. Am*, *103*(2), 1080-1084.

Darwin, C. J., & Hukin, R. W. (1999). Auditory objects of attention: The role of interaural time differences. *Journal of Experimental Psychology: Human Perception and Performance*, *25 (3)*, 617-629.

Darwin, C. J., & Hukin, R. W. (2000a). Effectiveness of spatial cues, prosody, and talker characteristics in selective attention. *J. Acoust. Soc. Am*, *107*(2), 970-977. doi: http://dx.doi.org/10.1121/1.428278

Darwin, C. J., & Hukin, R. W. (2000b). Effects of reverberation on spatial, prosodic, and vocal-tract size cues to selective attention. *J. Acoust. Soc. Am*, *108*(1), 335-342.

Demany, L. (1985). Perceptual learning in frequency discrimination. *J. Acoust. Soc. Am*, *78*(3), 1118-1120.

Devore, S., Ihlefeld, A., Hancock, K., Shinn-Cunningham, B., & Delgutte, B. (2009). Accurate sound localization in reverberant environments is mediated by robust encoding of spatial cues in the auditory midbrain. *Neuron*, *62 (1)*, 123 - 134.

Djelani, T., & Blauert, J. (2001). Investigations into the build-up and breakdown of the precedence effect. *Acta Acustica united with Acustica*, *87 (2)*, 253-261.

Durlach, N. I. (1963). Equalization and cancellation theory of binaural masking-level differences. *J. Acoust. Soc. Am*, *35 (8)*, 1206-1218.

Faller, C., & Merimaa, J. (2004). Source localization in complex listening situations: Selection of binaural cues based on interaural coherence. *J. Acoust. Soc. Am*, *116*(5), 3075–3089. doi: 10.1121/1.1791872

Festen, J. M., & Plomp, R. (1990). Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *J. Acoust. Soc. Am*, *88 (4)*, 1725-1736.

French, N., & Steinberg, J. C. (1947). Factors governing the intelligibility of speech sounds. *J. Acoust. Soc. Am*, *19 (1)*, 90-119.

Freyman, R. L., Clifton, R. K., & Litovsky, R. Y. (1991). Dynamic processes in the precedence effect. *J. Acoust. Soc. Am*, *90 (2)*, 874-884.

Freyman, R. L., Helfer, K. S., McCall, D. D., & Clifton, R. K. (1999). The role of perceived spatial separation in the unmasking of speech. *J. Acoust. Soc. Am*, *106*(6), 3578-3588.

Freyman, R. L., McCall, D. D., & Clifton, R. K. (1998). Intensity discrimination for precedence effect. *J. Acoust. Soc. Am*, *103 (4)*, 2031-2041.

Freyman, R. L., Zurek, P. M., Balakrishnan, U., & Chiang, Y.-C. (1997). Onset dominance in lateralization. *J. Acoust. Soc. Am*, *101*(3), 1649-1659.

Giguère, C., & Abel, S. M. (1993). Sound localization: Effects of reverberation time, speaker array, stimulus frequency, and stimulus rise/decay. *J. Acoust. Soc. Am*, *94*(2), 769-776.

Giguère, C., Lavallée, R., Plourde, J., & Vaillancourt, V. (2011). Vertical sound localization in left, median and right lateral planes. *Canadian Acoustics*, *39 (4)*, 3-12.

Gilbert, J. L., Tamati, T. N., & Pisoni, D. B. (2013, January). Development, reliability, and validity of presto: a new high-variability sentence recognition test. *Journal of the American Academy of Audiology*, *24*(1), 2636. doi: 10.3766/jaaa.24.1.4

Grantham, D. W. (1996). Left-right asymmetry in the buildup of echo suppression in normal-hearing adults. *J. Acoust. Soc. Am*, *99 (2)*, 1118-1123.

Green, D. M. (1961). Detection of auditory sinusoids of uncertain frequency. *The Journal of the Acoustical Society of America*, *33*(7), 897-903.

Hartmann, W. M. (1983). Localization of sound in rooms. *J. Acoust. Soc. Am*, *74*(5), 1380-1391.

Hartmann, W. M., & Rakerd, B. (1989). Localization of sound in rooms iv: The franssen effect. *J. Acoust. Soc. Am*, *86 (4)*, 1366-1373.

Hawley, M. L., Litovsky, R. Y., & Culling, J. F. (2004). The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer. *J. Acoust. Soc. Am*, *115*(2), 833-843.

Hirsh, I. J. (1950). The relation between localization and intelligibility. *J. Acoust. Soc. Am*, *22 (2)*, 196-200.

Ihlefeld, A., & Shinn-Cunningham, B. (2008). Spatial release from energetic and informational masking in a selective speech identification taska). *The Journal of the Acoustical Society of America*, *123*(6), 4369-4379.

Ihlefeld, A., & Shinn-Cunningham, B. G. (2011). Effect of source spectrum on sound localization in an everyday reverberant room. *J. Acoust. Soc. Am*, *130*(1), 324-333.

Jelfs, S., Culling, J. F., & Lavandier, M. (2011). Revision and validation of a binaural model for speech intelligibility in noise. *Hearing Research*, *275*, 96-104.

Jeub, M., Schafer, M., & Vary, P. (2009, July). A binaural room impulse response database for the evaluation of dereverberation algorithms. *16th International Conference on Digital Signal Processing*, 1-5. doi: 10.1109/ICDSP.2009.5201259

Kahneman, D. (1973). *Attention and effort.* Prentice-Hall.

Kidd, G., Arbogast, T. L., Mason, C. R., & Gallun, F. J. (2005). The advantage of knowing where to listen. *J. Acoust. Soc. Am*, *117*(4), 2537-2537. doi:

http://dx.doi.org/10.1121/1.4788428

Kidd, G., Mason, C. R., & Brughera, A. (2005). The role of reverberation in release from masking due to spatial separation of sources for speech identification. *Acta Acustica united with Acustica*, *91*, 526-536.

Kidd, G., Mason, C. R., Brughera, A., & Hartmann, W. M. (2005). The role of reverberation in release from masking due to spatial separation of sources for speech identification. *Acta Acustica united with Acustica*, *114*, 526-536.

Kryter, K. D. (1962). Validation of the articulation index. *J. Acoust. Soc. Am*, *34 (11)*, 1698-1702.

Kuttruff, H. (2000). *Room acoustics, 4th ed.* SPON press.

Lavandier, M., & Culling, J. F. (2010). Prediction of binaural speech intelligibility against noise in rooms. *The Journal of the Acoustical Society of America*, *127*(1), 387-399.

Lee, A. K. C., & Shinn-Cunningham, B. G. (2008). Effects of reverberant spatial cues on attention-dependent object formation. *Journal of the Association for Research in Otolaryngology*, *9*, 150-160.

Le Goff, N., Buchholz, J. M., & Dau, T. (2013). Spectral integration of interaural time differences in auditory localization. *Proceedings of Meetings on Acoustics*, *19*(1), 4aPPb22.

Litovsky, R. Y., Colburn, H. S., Yost, W. A., & Guzman, S. J. (1999). The precedence effect. *J. Acoust. Soc. Am*, *106(4)*, 1633.

Litovsky, R. Y., & Shinn-Cunningham, B. G. (2001). Investigation of the relationship among three common measures of precedence: Fusion, localization dominance, and discrimination suppression. *J. Acoust. Soc. Am*, *109*(1), 346-358.

Litovsky, R. Y., & Yin, T. C. T. (1998). Physiological studies of the precedence effect in the inferior colliculus of the cat. i. correlates of psychophysics. *Journal of Neurophysiology*, *80*, 1285-1301.

London, S., Bishop, C. W., & Miller, L. M. (2012). Spatial attention modulates the precedence effect. *Journal of Experimental Psychology: Human Perception and Performance*, *38 (6)*, 1371-1379.

Lutfi, R. A., Chang, A.-C., Stamas, J., & Gilbertson, L. (2012). A detection-theoretic framework for modeling informational masking. *The Journal of the Acoustical Society of America*, *132*(2), EL109-EL113.

Macpherson, E. A. (1995). Source spectrum recovery at different spatial locations. *The Journal of the Acoustical Society of America*, *98*(5), 2946-2946. doi: http://dx.doi.org/10.1121/1.414076

Macpherson, E. A., & Middlebrooks, J. C. (2002). Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited. *J. Acoust. Soc. Am*, *111*(5), 2219-2236.

Macpherson, E. A., & Sabin, A. T. (2007). Binaural weighting of monaural spectral cues for sound localization. , *121*(6), 3677-3688. doi: http://dx.doi.org/10.1121/1.2722048

Marrone, N., Mason, C. R., & Kidd., G. (2008). Tuning in the spatial dimension: Evidence from a masked speech identification task. *J. Acoust. Soc. Am*, *124*(2), 1146-1158.

McFadden, D., & Pasanen, E. G. (1976). Lateralization at high frequencies based on interaural time differences. *J. Acoust. Soc. Am*, *59*(3), 634-639.

Mendonça, C., Campos, G., Dias, P., & Santos, J. A. (2013). Learning auditory space: Generalization and long-term effects. *PLoS ONE*, *8(10)*, 1-14.

Nábělek, A. K., Letowski, T. R., & Tucker, F. M. (1989). Reverberant overlap- and self-masking in consonant identification. *J. Acoust. Soc. Am*, *86 (4)*, 1259-1265.

Nábělek, A. K., & Robinette, L. (1978). Influence of the precedence effect on word identification by normally hearing and hearing-impaired subjects. *J. Acoust. Soc. Am*, *63*(1), 187-194.

Nahum, M., Nelken, I., & Ahissar, M. (2010). Stimulus uncertainty and perceptual learning: Similar principles govern auditory and visual learning. *Vision Research*, *50*, 391-401.

Nguyen, T. M. (2014). *The effects of target spectrum, noise, and reverberation on auditory cue weighting in sound localization.* Unpublished master's thesis, University of Western Ontario.

Rakerd, B., & Hartmann, W. M. (1985). Localization of sound in rooms, ii: The effects of a single reflecting surface. *J. Acoust. Soc. Am*, *78(2)*, 524-533.

Rakerd, B., & Hartmann, W. M. (2010). Localization of sound in rooms. v. binaural coherence and human sensitivity to interaural time differences in noise. *J. Acoust. Soc. Am*, *128*(5), 3052-3063.

Rakerd, B., Hartmann, W. M., & McCaskey, T. L. (1999). Identification and localization of sound sources in the median sagittal plane. *J. Acoust. Soc. Am*, *106(5)*, 2812-2820.

Rife, D., & Vanderkooy, J. (1989). Transfer-function measurement with maximum-length sequences. *Journal of the Audio Engineering Society*, *37*, 419-444.

Schoolmaster, M., Kopco, N., & Shinn-Cunningham, B. G. (2003). Effects of reverberation and experience on distance perception in simulated environments. *J. Acoust. Soc. Am*, *113*(4), 2285-2285. doi: http://dx.doi.org/10.1121/1.4780593

Schroeder, M. R. (1965). New method of measuring reverberation time. *J. Acoust. Soc. Am*, *37*, 409-412.

Shinn-Cunningham, B. (2000). Learning reverberation: Considerations for spatial auditory display. *Proceedings of the 6th International Conference on Auditory Display (ICAD2000), 2-5 April, Atlanta, GA.* Retrieved from `http://hdl.handle.net/1853/50675`

Shinn-Cunningham, B., Kopčo, N., & Martin, T. J. (2005). Localizing nearby sound sources in a classroom: Binaural room impulse responses. *J. Acoust. Soc. Am*, *117 (5)*, 3100-3115.

Shinn-Cunningham, B., Zurek, P. M., & Durlach, N. I. (1993). Adjustment and discrimination measurements of the precedence effect. *J. Acoust. Soc. Am*, *93*(5), 2923-2932.

Spence, C. J., & Driver, J. (1994). Covert spatial orienting in audition: Exogenous and endogenous mechanisms. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 555.

Srinivasan, N. K., & Zahorik, P. (2013). Prior listening exposure to a reverberant room improves open-set intelligibility of high-variability sentences. *JASA Express Letters*, *133 (1)*, EL33-39.

Srinivasan, N. K., & Zahorik, P. (2014). Enhancement of speech intelligibility in reverberant rooms: Role of amplitude envelope and temporal fine structure.

The Journal of the Acoustical Society of America, *135*(6), EL239-EL245.

Steeneken, H., & Houtgast, T. (1980). A physical method for measuring speech-transmission quality. *J. Acoust. Soc. Am*, *67 (1)*, 318-326.

Stern, R. M., Zeiberg, A. S., & Trahiotis, C. (1988). Lateralization of complex binaural stimuli: A weighted-image model. *J. Acoust. Soc. Am*, *84 (1)*, 156-165.

Stevens, S. S., & Newman, E. B. (1936). The localization of actual sources of sound. *The American Journal of Psychology*, *48 (2)*, 297-306.

Strutt, J. W. (1907). On our perception of sound direction. *Philos. Mag.*, *13*, 214-232.

Studebaker, G. A. (1985). A “rationalized” arcsine transform. *Journal of Speech, Language, and Hearing Research*, *28*, 455-462.

Trahiotis, C., & Bernstein, L. R. (1986). Lateralization of bands of noise and sinusoidally amplitude-modulated tones: Effects of spectral locus and bandwidth. *J. Acoust. Soc. Am*, *79*, 19501957.

Wallach, H., Newman, E. B., & Rosenzweig, M. R. (1949). The precedence effect in sound localization. *The American Journal of Psychology*, *62*(3), pp. 315-336.

Wanrooij, M. M. V., & Opstal, A. J. V. (2005). Relearning sound localization with a new ear. *The Journal of Neuroscience*, *25(22)*, 5413-5424.

Watkins, A. J. (1991). Central, auditory mechanisms of perceptual compensation for spectral-envelope distortion. *J. Acoust. Soc. Am*, *90 (6)*, 2942-2955.

Watkins, A. J. (2005a). Perceptual compensation for effects of echo and of reverberation on speech identification. *Acta Acustica united with Acustica*, *91*, 892-901.

Watkins, A. J. (2005b). Perceptual compensation for effects of reverberation in speech identifications. *J. Acoust. Soc. Am*, *118 (1)*, 249-262.

Watkins, A. J., & Makin, S. J. (2007). Steady-spectrum contexts and perceptual compensation for reverberation in speech identification. *J. Acoust. Soc. Am*, *121*(1), 257-266.

Watkins, A. J., & Raimond, A. P. (2013). Basic aspects of hearing: Physiology and perception. In B. C. J. Moore, R. D. Patterson, I. M. Winter, R. P. Carlyon, & H. E. Gockel (Eds.), (p. 193-201). Springer.

Watkins, A. J., Raimond, A. P., & Makin, S. J. (2011). Temporal-envelope constancy of speech in rooms and the perceptual weighting of frequency bands. *J. Acoust. Soc. Am*, *130 (5)*, 2777-2788.

Watson, C. S., Kelly, W. J., & Wroton, H. W. (1976). Factors in the discrimination of tonal patterns. ii. selective attention and learning under various levels of stimulus uncertainty. *J. Acoust. Soc. Am*, *60*(5), 1176-1186.

Wightman, F. L., & Kistler, D. J. (1992). The dominant role of low-frequency interaural time differences in sound localization. *J. Acoust. Soc. Am*, *91*, 1648-1661.

Wright, B. A., & Fitzgerald, M. B. (2001). Different patterns of human discrimination learning for two interaural cues to sound-source location. *Proceedings of the National Academy of Sciences*, *98*(21), 12307-12312. doi: 10.1073/pnas.211220498

Yang, X., & Grantham, D. W. (1997). Echo suppression and discrimination suppression aspects of the precedence effect. *Perception and Psychophysics*, *59 (7)*, 1108-1117.

Yost, W. A., Loiselle, L., Dorman, M., Burns, J., & Brown, C. A. (2013). Sound source localization of filtered noises by listeners with normal hearing: A statistical analysis. *J. Acoust. Soc. Am*, *133*(5), 2876-2882.

Zahorik, P. (2009). Perceptually relevant parameters for virtual listening simulation of small room acoustics. *J. Acoust. Soc. Am*, *126*(2), 776-791. doi: http://dx.doi.org/10.1121/1.3167842

Zahorik, P., & Brandewie, E. (2011). Perceptual adaptation to room acoustics and effects on speech intelligibility in hearing-impaired populations. *Proc. Forum Acust.*, *27*, 2167-2172.

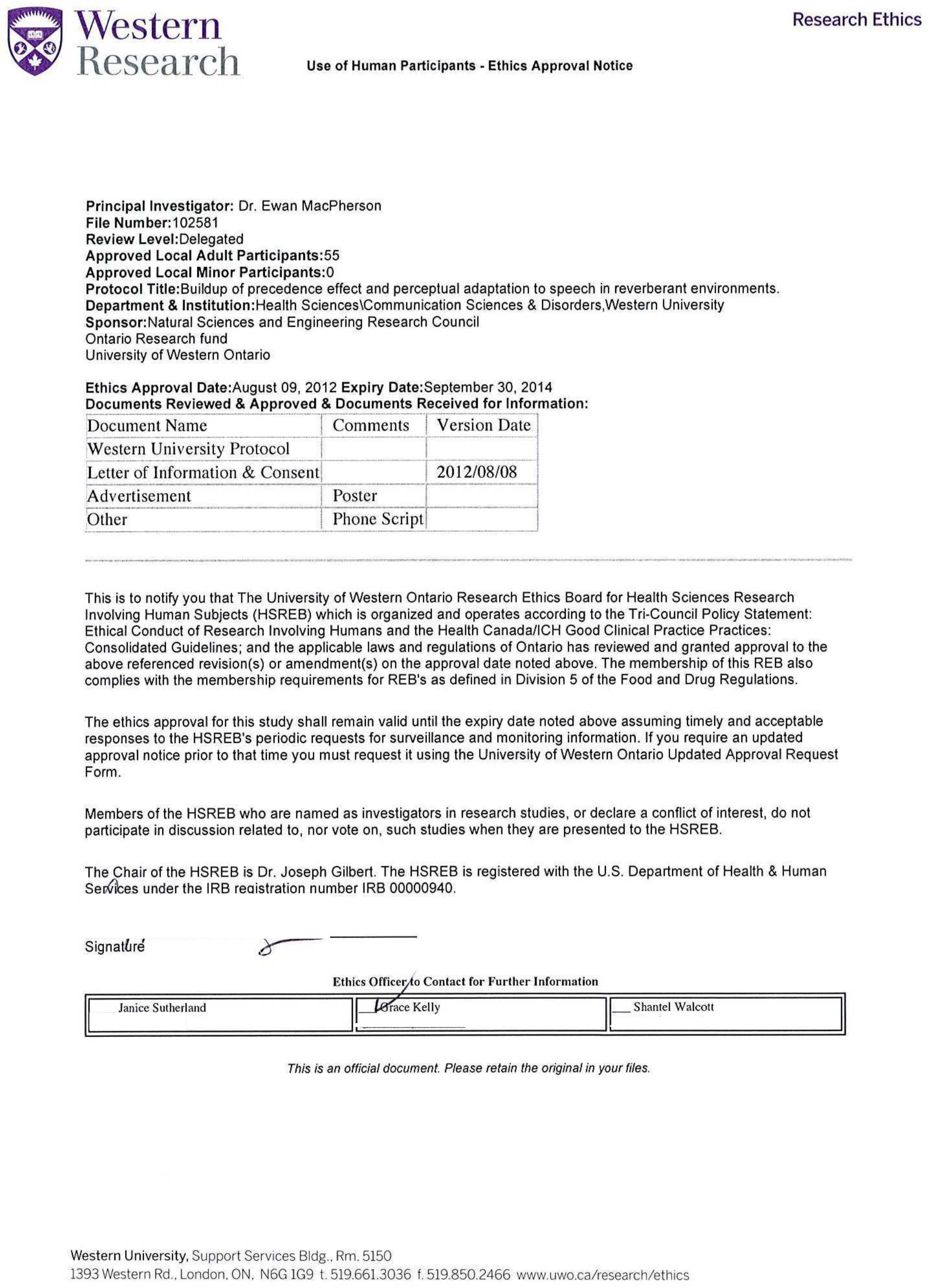
Zahorik, P., Brandewie, E., & Sivonen, V. P. (2009). Spatial hearing in reverberant rooms and effects of prior listening exposure. *International Workshop on the Principles and Applications of Spatial Hearing, November 11-13, Zao, Miyagi, Japan.*

Zurek, P. (1980). The precedence effect and its possible role in the avoidance of interaural ambiguities. *J. Acoust. Soc. Am*, *67 (3)*, 952-964.

Zwislocki, J., & Feldman, R. S. (1956). Just noticeable differences in dichotic phase. *J. Acoust. Soc. Am*, *28*, 860864.

Appendix A

Ethics Approval



Western Research

Research Ethics

Use of Human Participants - Ethics Approval Notice

Principal Investigator: Dr. Ewan MacPherson
File Number:102581
Review Level: Delegated
Approved Local Adult Participants:55
Approved Local Minor Participants:0
Protocol Title: Buildup of precedence effect and perceptual adaptation to speech in reverberant environments.
Department & Institution: Health Sciences\Communication Sciences & Disorders,Western University
Sponsor: Natural Sciences and Engineering Research Council
Ontario Research fund
University of Western Ontario

Ethics Approval Date: August 09, 2012 **Expiry Date:**September 30, 2014
Documents Reviewed & Approved & Documents Received for Information:

Document Name	Comments	Version Date
Western University Protocol		
Letter of Information & Consent		2012/08/08
Advertisement	Poster	
Other	Phone Script	

This is to notify you that The University of Western Ontario Research Ethics Board for Health Sciences Research Involving Human Subjects (HSREB) which is organized and operates according to the Tri-Council Policy Statement: Ethical Conduct of Research Involving Humans and the Health Canada/ICH Good Clinical Practice Practices: Consolidated Guidelines; and the applicable laws and regulations of Ontario has reviewed and granted approval to the above referenced revision(s) or amendment(s) on the approval date noted above. The membership of this REB also complies with the membership requirements for REB's as defined in Division 5 of the Food and Drug Regulations.

The ethics approval for this study shall remain valid until the expiry date noted above assuming timely and acceptable responses to the HSREB's periodic requests for surveillance and monitoring information. If you require an updated approval notice prior to that time you must request it using the University of Western Ontario Updated Approval Request Form.

Members of the HSREB who are named as investigators in research studies, or declare a conflict of interest, do not participate in discussion related to, nor vote on, such studies when they are presented to the HSREB.

The Chair of the HSREB is Dr. Joseph Gilbert. The HSREB is registered with the U.S. Department of Health & Human Services under the IRB registration number IRB 00000940.

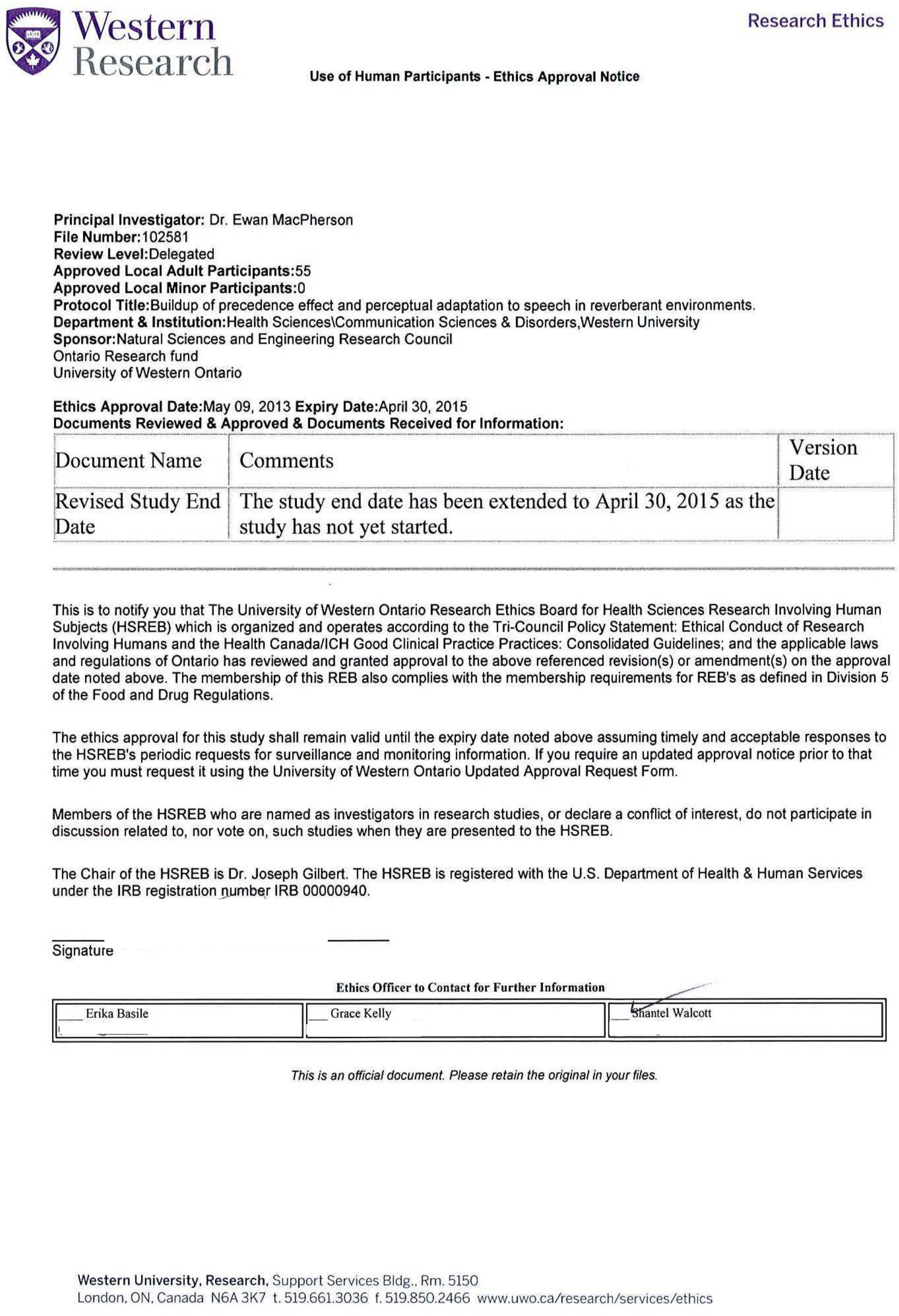
Signature

Ethics Officer to Contact for Further Information

Janice Sutherland	✓ Grace Kelly	Shantel Walcott

This is an official document. Please retain the original in your files.

Western University, Support Services Bldg., Rm. 5150
1393 Western Rd., London, ON, N6G 1G9 t. 519.661.3036 f. 519.850.2466 www.uwo.ca/research/ethics



Western Research

Research Ethics

Use of Human Participants - Ethics Approval Notice

Principal Investigator: Dr. Ewan MacPherson
File Number:102581
Review Level: Delegated
Approved Local Adult Participants:55
Approved Local Minor Participants:0
Protocol Title: Buildup of precedence effect and perceptual adaptation to speech in reverberant environments.
Department & Institution: Health Sciences\Communication Sciences & Disorders,Western University
Sponsor: Natural Sciences and Engineering Research Council
Ontario Research fund
University of Western Ontario

Ethics Approval Date: May 09, 2013 **Expiry Date:**April 30, 2015
Documents Reviewed & Approved & Documents Received for Information:

Document Name	Comments	Version Date
Revised Study End Date	The study end date has been extended to April 30, 2015 as the study has not yet started.	

This is to notify you that The University of Western Ontario Research Ethics Board for Health Sciences Research Involving Human Subjects (HSREB) which is organized and operates according to the Tri-Council Policy Statement: Ethical Conduct of Research Involving Humans and the Health Canada/ICH Good Clinical Practice Practices: Consolidated Guidelines; and the applicable laws and regulations of Ontario has reviewed and granted approval to the above referenced revision(s) or amendment(s) on the approval date noted above. The membership of this REB also complies with the membership requirements for REB's as defined in Division 5 of the Food and Drug Regulations.

The ethics approval for this study shall remain valid until the expiry date noted above assuming timely and acceptable responses to the HSREB's periodic requests for surveillance and monitoring information. If you require an updated approval notice prior to that time you must request it using the University of Western Ontario Updated Approval Request Form.

Members of the HSREB who are named as investigators in research studies, or declare a conflict of interest, do not participate in discussion related to, nor vote on, such studies when they are presented to the HSREB.

The Chair of the HSREB is Dr. Joseph Gilbert. The HSREB is registered with the U.S. Department of Health & Human Services under the IRB registration number IRB 00000940.

Signature

Ethics Officer to Contact for Further Information

Erika Basile	Grace Kelly	Shantel Walcott

This is an official document. Please retain the original in your files.

Western University, Research, Support Services Bldg., Rm. 5150
London, ON, Canada N6A 3K7 t. 519.661.3036 f. 519.850.2466 www.uwo.ca/research/services/ethics

Appendix B

Estimated $P(x)$ parameters for the speech-in-noise perception experiment

Table B.1: Estimated α, β in eqn. (5.1), their corresponding 95% confidence intervals, and the goodness-of-fit, R^2, for all listening conditions in R0.

SID	Mixed-Short			Fixed-Short			Fixed-Full		
	$\hat{\alpha}$ (±95% CI)	$\hat{\beta}$ (±95% CI)	R^2	$\hat{\alpha}$ (±95% CI)	$\hat{\beta}$ (±95% CI)	R^2	$\hat{\alpha}$ (±95% CI)	$\hat{\beta}$ (±95% CI)	R^2
S26	-15.31 (±3.24)	0.31 (±0.33)	0.953	-16.98 (±2.17)	0.26 (±0.16)	0.980	-18.15 (±2.55)	0.25 (±0.16)	0.980
S27	-15.63 (±1.31)	0.46 (±0.26)	0.993	-17.03 (±2.46)	0.63 (±0.95)	0.951	-18.66 (±0.91)	0.33 (±0.10)	0.996
S28	-15.62 (±1.88)	0.31 (±0.19)	0.986	-15.63 (±2.93)	0.39 (±0.44)	0.959	-15.20 (±5.22)	0.33 (±0.59)	0.882
S29	-16.67 (±1.46)	0.44 (±0.29)	0.986	-19.30 (±8.24)	0.22 (±0.39)	0.846	-18.56 (±2.79)	0.36 (±0.40)	0.957
S30	-19.61 (±4.80)	0.26 (±0.30)	0.948	-19.07 (±1.46)	0.34 (±0.17)	0.990	-19.71 (±0.69)	0.47 (±0.14)	0.998
S31	-17.05 (±0.19)	0.51 (±0.05)	1.000	-18.14 (±1.06)	0.35 (±0.14)	0.995	-21.50 (±1.63)	0.23 (±0.07)	0.996
S32	-18.83 (±2.96)	0.32 (±0.30)	0.964	-16.29 (±0.44)	0.29 (±0.04)	0.999	-18.17 (±0.41)	0.39 (±0.07)	0.999
S18	-19.15 (±2.87)	0.26 (±0.19)	0.979	-18.66 (±0.91)	0.33 (±0.10)	0.996	-18.60 (±3.83)	0.25 (±0.24)	0.952
S33	-17.16 (±2.36)	0.32 (±0.27)	0.971	-17.69 (±0.68)	0.28 (±0.06)	0.998	-16.11 (±1.65)	0.38 (±0.25)	0.984
S34	-15.03 (±0.44)	0.31 (±0.04)	0.999	-16.44 (±0.81)	0.42 (±0.15)	0.996	-16.81 (±0.48)	0.30 (±0.05)	0.999
Mean (SD)	-17.00 (1.68)	0.35	0.978	-17.52 (1.25)	0.35	0.971	-18.15 (1.79)	0.33	0.974
Median	-16.86	0.32	0.982	-17.36	0.33	0.993	-18.36	0.33	0.990
Min	-19.61	0.26	0.948	-19.30	0.22	0.846	-21.50	0.23	0.882
Max	-15.03	0.51	1.000	-15.63	0.63	0.999	-15.20	0.47	0.999

Table B.2: Estimated α, β in eqn. (5.1), their corresponding 95% confidence intervals, and the goodness-of-fit, R^2, for all listening conditions in R1.

SID	Mixed-Short			Fixed-Short			Fixed-Full		
	$\hat{\alpha}$ (±95% CI)	$\hat{\beta}$ (±95% CI)	R^2	$\hat{\alpha}$ (±95% CI)	$\hat{\beta}$ (±95% CI)	R^2	$\hat{\alpha}$ (±95% CI)	$\hat{\beta}$ (±95% CI)	R^2
S26	-16.29 (±2.55)	0.38 (±0.38)	0.963	-14.77 (±5.63)	0.28 (±0.47)	0.877	-16.08 (±2.06)	0.45 (±0.40)	0.975
S27	-16.80 (±1.18)	0.25 (±0.08)	0.994	-15.54 (±3.80)	0.29 (±0.35)	0.938	-18.21 (±6.14)	0.19 (±0.24)	0.899
S28	-15.21 (±2.52)	0.33 (±0.29)	0.972	-14.91 (±1.39)	0.45 (±0.26)	0.991	-17.74 (±0.90)	0.33 (±0.10)	0.997
S29	-16.69 (±1.60)	0.46 (±0.34)	0.989	-17.93 (±2.35)	0.47 (±0.60)	0.961	-26.19 (±22.18)	0.14 (±0.26)	0.745
S30	-19.93 (±0.67)	0.26 (±0.04)	0.999	-19.17 (±2.83)	0.44 (±0.58)	0.966	-19.80 (±2.55)	0.26 (±0.15)	0.986
S31	-17.89 (±1.14)	0.44 (±0.25)	0.990	-17.12 (±2.24)	0.29 (±0.21)	0.977	-17.66 (±1.45)	0.57 (±0.57)	0.986
S32	-17.03 (±2.87)	0.24 (±0.18)	0.967	-17.02 (±1.82)	0.45 (±0.38)	0.980	-17.04 (±5.11)	0.26 (±0.37)	0.905
S18	-18.82 (±0.53)	0.24 (±0.03)	0.999	-22.99 (±3.45)	0.19 (±0.09)	0.988	-19.61 (±7.78)	0.22 (±0.35)	0.889
S33	-16.47 (±1.03)	0.29 (±0.09)	0.995	-15.82 (±2.81)	0.36 (±0.37)	0.961	-17.04 (±1.12)	0.48 (±0.27)	0.994
S34	-16.33 (±4.07)	0.27 (±0.32)	0.927	-18.60 (±11.41)	0.15 (±0.29)	0.784	-17.75 (±0.90)	0.25 (±0.06)	0.997
Mean (SD)	-17.15 (1.38)	0.32	0.979	-17.39 (2.48)	0.34	0.942	-18.71 (2.86)	0.31	0.937
Median	-16.75	0.28	0.989	-17.07	0.33	0.964	-17.74	0.26	0.980
Min	-19.93	0.24	0.927	-22.99	0.15	0.784	-26.19	0.14	0.745
Max	-15.21	0.46	0.999	-14.77	0.47	0.991	-16.08	0.57	0.997

Table B.3: Estimated α, β in eqn. (5.1), their corresponding 95% confidence intervals, and the goodness-of-fit, R^2, for all listening conditions in R2.

SID	Mixed-Short			Fixed-Short			Fixed-Full		
	$\hat{\alpha}$ (±95% CI)	$\hat{\beta}$ (±95% CI)	R^2	$\hat{\alpha}$ (±95% CI)	$\hat{\beta}$ (±95% CI)	R^2	$\hat{\alpha}$ (±95% CI)	$\hat{\beta}$ (±95% CI)	R^2
S26	-15.71 (±2.29)	0.43 (±0.40)	0.977	-16.65 (±3.38)	0.22 (±0.19)	0.955	-16.13 (±1.76)	0.39 (±0.27)	0.985
S27	-17.64 (±13.02)	0.17 (±0.45)	0.621	-17.09 (±6.84)	0.27 (±0.54)	0.813	-19.09 (±5.49)	0.22 (±0.27)	0.927
S28	-16.98 (±2.68)	0.17 (±0.09)	0.976	-15.82 (±3.50)	0.23 (±0.21)	0.953	-15.99 (±0.65)	0.46 (±0.13)	0.998
S29	-16.68 (±4.13)	0.23 (±0.25)	0.931	-17.05 (±2.20)	0.65 (±0.90)	0.976	-21.14 (±4.48)	0.19 (±0.13)	0.972
S30	-16.98 (±2.67)	0.30 (±0.27)	0.971	-18.94 (±3.06)	0.20 (±0.12)	0.978	-17.66 (±0.20)	0.60 (±0.09)	1.000
S31	-17.50 (±2.48)	0.29 (±0.23)	0.976	-17.92 (±2.60)	0.30 (±0.25)	0.969	-20.58 (±3.55)	0.21 (±0.14)	0.976
S32	-17.13 (±7.00)	0.26 (±0.54)	0.814	-16.42 (±1.10)	0.22 (±0.06)	0.995	-18.20 (±5.56)	0.28 (±0.45)	0.880
S18	-18.07 (±4.97)	0.21 (±0.23)	0.935	-20.57 (±9.61)	0.14 (±0.18)	0.885	-19.05 (±2.60)	0.24 (±0.14)	0.980
S33	-18.72 (±5.33)	0.21 (±0.24)	0.919	-16.10 (±1.00)	0.27 (±0.08)	0.995	-18.00 (±2.81)	0.31 (±0.28)	0.970
S34	-15.60 (±2.07)	0.25 (±0.15)	0.981	-15.01 (±2.01)	0.33 (±0.22)	0.983	-15.91 (±2.65)	0.32 (±0.29)	0.963
Mean (SD)	-17.10 (0.97)	0.25	0.910	-17.16 (1.63)	0.28	0.950	-18.18 (1.84)	0.32	0.965
Median	-17.05	0.24	0.953	-16.85	0.25	0.972	-18.10	0.29	0.974
Min	-18.72	0.17	0.621	-20.57	0.14	0.813	-21.14	0.19	0.880
Max	-15.60	0.43	0.981	-15.01	0.65	0.995	-15.91	0.60	1.000

Table B.4: Estimated α, β in eqn. (5.1), their corresponding 95% confidence intervals, and the goodness-of-fit, R^2, for all listening conditions in R3.

SID	Mixed-Short			Fixed-Short			Fixed-Full		
	$\hat{\alpha}$ (±95% CI)	$\hat{\beta}$ (±95% CI)	R^2	$\hat{\alpha}$ (±95% CI)	$\hat{\beta}$ (±95% CI)	R^2	$\hat{\alpha}$ (±95% CI)	$\hat{\beta}$ (±95% CI)	R^2
S26	-10.74 (±1.35)	0.35 (±0.16)	0.992	-12.13 (±3.09)	0.41 (±0.47)	0.961	-13.03 (±1.43)	0.36 (±0.18)	0.992
S27	-12.68 (±7.07)	0.25 (±0.50)	0.829	-12.99 (±2.53)	0.32 (±0.26)	0.975	-13.43 (±1.71)	0.33 (±0.18)	0.987
S28	-12.52 (±1.89)	0.30 (±0.17)	0.985	-12.89 (±1.56)	0.36 (±0.19)	0.991	-11.05 (±2.64)	0.43 (±0.45)	0.976
S29	-12.01 (±0.42)	0.50 (±0.09)	0.999	-10.66 (±1.86)	0.45 (±0.35)	0.988	-13.64 (±1.23)	0.32 (±0.13)	0.993
S30	-13.96 (±1.03)	0.22 (±0.06)	0.995	-13.96 (±1.03)	0.22 (±0.06)	0.995	-15.44 (±1.67)	0.33 (±0.18)	0.987
S31	-12.72 (±0.43)	0.36 (±0.05)	0.999	-12.59 (±1.68)	0.32 (±0.17)	0.989	-14.13 (±1.37)	0.38 (±0.19)	0.991
S32	-12.68 (±1.65)	0.37 (±0.21)	0.990	-11.57 (±2.63)	0.80 (±1.11)	0.983	-14.53 (±3.18)	0.42 (±0.54)	0.951
S18	-12.20 (±0.22)	0.36 (±0.03)	1.000	-13.30 (±4.76)	0.32 (±0.48)	0.915	-13.53 (±2.29)	0.30 (±0.21)	0.980
S33	-12.10 (±2.69)	0.34 (±0.30)	0.969	-11.15 (±2.19)	0.37 (±0.28)	0.980	-15.12 (±2.35)	0.54 (±0.61)	0.980
S34	-12.02 (±4.63)	0.27 (±0.35)	0.914	-11.89 (±0.77)	0.46 (±0.14)	0.998	-12.92 (±2.71)	0.27 (±0.21)	0.970
Mean (SD)	-12.36 (0.81)	0.33	0.967	-12.31 (1.02)	0.40	0.978	-13.68 (1.25)	0.37	0.981
Median	-12.36	0.35	0.991	-12.36	0.36	0.986	-13.59	0.35	0.984
Min	-13.96	0.22	0.829	-13.96	0.22	0.915	-15.44	0.27	0.951
Max	-10.74	0.50	1.000	-10.66	0.80	0.998	-11.05	0.54	0.993

Appendix C

Estimated $P(x)$ parameters for the multi-talker speech perception experiment

Table C.1: Estimated α, β in eqn. (5.1), their corresponding 95% confidence intervals, and goodness-of-fit, R^2, for all listening conditions in R0.

(a) Mixed condition

SID	Co-located			Spatially-separated		
	$\hat{\alpha}$ (±95% CI)	$\hat{\beta}$ (±95% CI)	R^2	$\hat{\alpha}$ (±95% CI)	$\hat{\beta}$ (±95% CI)	R^2
S03	3.8 (±0.4)	0.6 (±0.2)	0.99	0.7 (±1.7)	0.2 (±0.1)	0.90
S08	4.7 (±0.8)	0.5 (±0.2)	0.98	4.3 (±0.5)	0.5 (±0.1)	0.99
S11	4.2 (±1.8)	0.4 (±0.3)	0.90	1.1 (±1.6)	0.4 (±0.3)	0.92
S12	2.8 (±1.4)	0.4 (±0.2)	0.94	-0.8 (±0.8)	0.3 (±0.1)	0.98
S13	3.2 (±0.5)	0.6 (±0.2)	0.99	-1.1 (±0.8)	0.4 (±0.1)	0.98
S14	2.8 (±2.5)	0.4 (±0.4)	0.83	0.5 (±2.9)	0.2 (±0.1)	0.74
S15	3.1 (±1.3)	0.5 (±0.3)	0.95	-1.7 (±1.2)	0.3 (±0.2)	0.94
S16	3.0 (±1.9)	0.6 (±0.7)	0.88	-1.3 (±1.6)	0.4 (±0.3)	0.91
S17	1.9 (±1.1)	0.4 (±0.2)	0.96	-3.3 (±1.7)	0.4 (±0.3)	0.90
S18	2.3 (±1.1)	0.5 (±0.3)	0.96	-1.7 (±1.8)	0.4 (±0.2)	0.89
S19	2.6 (±0.8)	0.6 (±0.2)	0.98	-2.3 (±1.5)	0.3 (±0.1)	0.93
Mean (SD)	3.1 (0.8)	0.5 (0.1)	0.94	-0.5 (2.1)	0.3 (0.1)	0.92
Median	3.0	0.5	0.96	-1.1	0.4	0.92

(b) Fixed condition

SID	Co-located			Spatially-separated		
	$\hat{\alpha}$ (±95% CI)	$\hat{\beta}$ (±95% CI)	R^2	$\hat{\alpha}$ (±95% CI)	$\hat{\beta}$ (±95% CI)	R^2
S03	3.3 (±1.6)	0.4 (±0.2)	0.92	1.7 (±1.0)	0.4 (±0.1)	0.97
S08	4.1 (±2.6)	0.6 (±0.9)	0.85	1.3 (±1.4)	0.4 (±0.2)	0.93
S11	4.2 (±0.9)	0.5 (±0.2)	0.97	0.7 (±0.6)	0.3 (±0.1)	0.99
S12	2.3 (±0.5)	0.6 (±0.1)	0.99	-1.1 (±0.7)	0.4 (±0.1)	0.98
S13	3.1 (±0.4)	0.5 (±0.1)	0.99	-0.8 (±0.4)	0.3 (±0.1)	0.99
S14	4.0 (±0.8)	0.6 (±0.3)	0.98	0.4 (±2.2)	0.2 (±0.1)	0.84
S15	2.8 (±1.3)	0.5 (±0.3)	0.95	-2.4 (±1.4)	0.3 (±0.1)	0.94
S16	1.8 (±1.0)	0.7 (±0.4)	0.96	-4.9 (±4.9)	0.2 (±0.3)	0.72
S17	2.0 (±1.3)	0.3 (±0.1)	0.94	-3.0 (±0.5)	0.4 (±0.1)	0.99
S18	2.3 (±1.0)	0.4 (±0.2)	0.97	-3.1 (±1.9)	0.3 (±0.2)	0.89
S19	3.0 (±1.3)	0.5 (±0.3)	0.94	-3.7 (±3.2)	0.3 (±0.2)	0.81
Mean (SD)	3.0 (0.8)	0.5 (0.1)	0.95	-1.4 (2.2)	0.3 (0.1)	0.91
Median	3.0	0.5	0.96	-1.1	0.3	0.94

Table C.2: Estimated α, β in eqn. (5.1), their corresponding 95% confidence intervals, and goodness-of-fit, R^2, for all listening conditions in R1.

(a) Mixed condition

SID	Co-located			Spatially-separated		
	$\hat{\alpha}$ (±95% CI)	$\hat{\beta}$ (±95% CI)	R^2	$\hat{\alpha}$ (±95% CI)	$\hat{\beta}$ (±95% CI)	R^2
S03	3.1 (±0.9)	0.6 (±0.3)	0.97	2.1 (±1.4)	0.4 (±0.3)	0.93
S08	5.0 (±0.8)	0.3 (±0.1)	0.98	4.9 (±1.9)	0.5 (±0.5)	0.90
S11	4.5 (±1.7)	0.4 (±0.3)	0.91	3.3 (±0.9)	0.5 (±0.2)	0.97
S12	2.8 (±0.4)	0.7 (±0.2)	1.00	1.4 (±0.8)	0.4 (±0.1)	0.98
S13	3.0 (±0.9)	0.6 (±0.3)	0.98	0.5 (±0.9)	0.3 (±0.1)	0.97
S14	3.6 (±0.4)	0.6 (±0.1)	0.99	2.6 (±1.1)	0.4 (±0.2)	0.96
S15	2.5 (±0.6)	0.6 (±0.2)	0.99	0.6 (±0.8)	0.3 (±0.1)	0.97
S16	3.8 (±1.9)	0.5 (±0.4)	0.90	-0.2 (±0.7)	0.4 (±0.1)	0.98
S17	2.6 (±0.9)	0.6 (±0.2)	0.98	-0.8 (±0.9)	0.4 (±0.2)	0.97
S18	3.2 (±1.1)	0.5 (±0.2)	0.96	1.2 (±0.8)	0.4 (±0.1)	0.98
S19	3.2 (±0.6)	0.6 (±0.2)	0.99	0.3 (±1.8)	0.4 (±0.3)	0.90
Mean (SD)	3.4 (0.8)	0.5 (0.1)	0.97	1.4 (1.7)	0.4 (0.0)	0.95
Median	3.2	0.6	0.98	1.2	0.4	0.97

(b) Fixed condition

SID	Co-located			Spatially-separated		
	$\hat{\alpha}$ (±95% CI)	$\hat{\beta}$ (±95% CI)	R^2	$\hat{\alpha}$ (±95% CI)	$\hat{\beta}$ (±95% CI)	R^2
S03	4.0 (±0.8)	0.5 (±0.2)	0.98	1.7 (±1.3)	0.4 (±0.2)	0.94
S08	4.5 (±1.4)	0.5 (±0.4)	0.95	3.2 (±2.1)	0.4 (±0.3)	0.87
S11	4.8 (±0.7)	0.5 (±0.2)	0.98	1.9 (±1.4)	0.3 (±0.2)	0.94
S12	3.6 (±0.5)	0.4 (±0.1)	0.99	-0.3 (±0.6)	0.4 (±0.1)	0.99
S13	4.1 (±0.9)	0.6 (±0.3)	0.97	0.8 (±0.5)	0.4 (±0.1)	0.99
S14	3.9 (±1.0)	0.7 (±0.4)	0.96	3.2 (±2.3)	0.2 (±0.1)	0.83
S15	3.4 (±1.1)	0.5 (±0.3)	0.96	0.4 (±1.4)	0.4 (±0.2)	0.93
S16	3.3 (±0.5)	0.8 (±0.3)	0.99	-2.2 (±0.9)	0.4 (±0.1)	0.96
S17	3.4 (±1.6)	0.5 (±0.4)	0.93	-1.0 (±1.1)	0.3 (±0.1)	0.96
S18	3.1 (±0.7)	0.7 (±0.3)	0.98	-0.3 (±3.5)	0.2 (±0.3)	0.68
S19	3.5 (±1.7)	0.4 (±0.3)	0.92	-0.9 (±0.6)	0.5 (±0.2)	0.99
Mean (SD)	3.8 (0.5)	0.6 (0.1)	0.96	0.6 (1.8)	0.3 (0.1)	0.92
Median	3.6	0.5	0.97	0.4	0.4	0.94

Table C.3: Estimated α, β in eqn. (5.1), their corresponding 95% confidence intervals, and goodness-of-fit, R^2, for all listening conditions in R2.

(a) Mixed condition

SID	Co-located			Spatially-separated		
	$\hat{\alpha}$ (±95% CI)	$\hat{\beta}$ (±95% CI)	R^2	$\hat{\alpha}$ (±95% CI)	$\hat{\beta}$ (±95% CI)	R^2
S03	2.8 (±0.3)	0.6 (±0.1)	1.00	4.0 (±1.4)	0.4 (±0.3)	0.93
S08	4.8 (±2.0)	0.3 (±0.3)	0.89	4.5 (±1.8)	0.5 (±0.4)	0.91
S11	4.5 (±1.7)	0.5 (±0.3)	0.90	3.7 (±0.5)	0.5 (±0.1)	0.99
S12	3.2 (±0.9)	0.5 (±0.2)	0.97	0.5 (±1.4)	0.6 (±0.4)	0.92
S13	3.7 (±1.2)	0.5 (±0.3)	0.95	1.3 (±1.1)	0.4 (±0.2)	0.95
S14	4.3 (±1.1)	0.5 (±0.3)	0.96	3.7 (±1.0)	0.4 (±0.2)	0.97
S15	2.9 (±0.6)	0.4 (±0.1)	0.99	-0.1 (±2.8)	0.3 (±0.3)	0.78
S16	3.0 (±0.4)	0.6 (±0.2)	0.99	-0.2 (±0.6)	0.4 (±0.1)	0.98
S17	3.1 (±1.2)	0.5 (±0.3)	0.95	-1.5 (±3.2)	0.3 (±0.3)	0.79
S18	3.7 (±1.0)	0.4 (±0.2)	0.96	0.5 (±1.9)	0.3 (±0.2)	0.88
S19	2.7 (±1.2)	0.4 (±0.2)	0.95	-1.0 (±2.7)	0.3 (±0.2)	0.83
Mean (SD)	3.5 (0.7)	0.5 (0.1)	0.96	1.4 (2.2)	0.4 (0.1)	0.90
Median	3.2	0.5	0.96	0.5	0.4	0.92

(b) Fixed condition

SID	Co-located			Spatially-separated		
	$\hat{\alpha}$ (±95% CI)	$\hat{\beta}$ (±95% CI)	R^2	$\hat{\alpha}$ (±95% CI)	$\hat{\beta}$ (±95% CI)	R^2
S03	3.3 (±0.4)	0.7 (±0.2)	1.00	3.8 (±0.9)	0.8 (±0.5)	0.98
S08	3.8 (±0.3)	0.6 (±0.1)	1.00	4.4 (±1.1)	0.4 (±0.2)	0.97
S11	4.6 (±1.2)	0.5 (±0.3)	0.95	1.8 (±2.9)	0.3 (±0.3)	0.76
S12	3.1 (±0.9)	0.6 (±0.3)	0.97	2.3 (±0.7)	0.5 (±0.2)	0.98
S13	4.1 (±1.1)	0.3 (±0.1)	0.96	1.5 (±0.5)	0.5 (±0.1)	0.99
S14	3.7 (±1.0)	0.5 (±0.2)	0.96	3.6 (±1.8)	0.5 (±0.4)	0.90
S15	2.9 (±0.7)	0.5 (±0.2)	0.98	1.2 (±2.2)	0.3 (±0.3)	0.84
S16	2.7 (±0.9)	0.4 (±0.1)	0.97	0.7 (±0.4)	0.7 (±0.2)	0.99
S17	4.2 (±0.7)	0.6 (±0.2)	0.98	-1.5 (±1.7)	0.3 (±0.2)	0.93
S18	2.6 (±0.2)	0.4 (±0.0)	1.00	0.5 (±1.4)	0.3 (±0.2)	0.93
S19	2.9 (±0.7)	0.4 (±0.1)	0.98	-1.0 (±1.1)	0.4 (±0.2)	0.95
Mean (SD)	3.4 (0.7)	0.5 (0.1)	0.98	1.6 (1.9)	0.5 (0.2)	0.93
Median	3.3	0.5	0.98	1.5	0.4	0.95

Table C.4: Estimated α, β in eqn. (5.1), their corresponding 95% confidence intervals, and goodness-of-fit, R^2, for all listening conditions in R3.

(a) Mixed condition

SID	Co-located			Spatially-separated		
	$\hat{\alpha}$ (±95% CI)	$\hat{\beta}$ (±95% CI)	R^2	$\hat{\alpha}$ (±95% CI)	$\hat{\beta}$ (±95% CI)	R^2
S03	4.2 (±0.6)	0.5 (±0.2)	0.99	4.2 (±1.8)	0.3 (±0.2)	0.89
S08	5.4 (±0.5)	0.5 (±0.1)	0.99	3.7 (±1.7)	0.6 (±0.6)	0.93
S11	4.9 (±2.1)	0.4 (±0.3)	0.87	2.5 (±2.0)	0.4 (±0.3)	0.86
S12	2.9 (±1.0)	0.4 (±0.1)	0.97	2.6 (±1.1)	0.6 (±0.3)	0.95
S13	3.6 (±1.3)	0.6 (±0.5)	0.94	3.2 (±0.6)	0.5 (±0.1)	0.99
S14	3.9 (±1.1)	0.5 (±0.3)	0.95	2.8 (±1.8)	0.5 (±0.4)	0.90
S15	3.2 (±0.5)	0.7 (±0.2)	0.99	1.3 (±0.8)	0.4 (±0.1)	0.97
S16	3.3 (±1.3)	0.4 (±0.2)	0.94	0.9 (±1.4)	0.4 (±0.2)	0.93
S17	2.8 (±1.6)	0.4 (±0.3)	0.91	1.5 (±1.3)	0.4 (±0.2)	0.94
S18	2.7 (±0.6)	0.4 (±0.1)	0.98	3.0 (±2.1)	0.4 (±0.4)	0.87
S19	3.2 (±0.4)	0.7 (±0.2)	0.99	1.8 (±1.3)	0.4 (±0.2)	0.95
Mean (SD)	3.6 (0.9)	0.5 (0.1)	0.96	2.5 (1.0)	0.4 (0.1)	0.93
Median	3.3	0.5	0.97	2.6	0.4	0.93

(b) Fixed condition

SID	Co-located			Spatially-separated		
	$\hat{\alpha}$ (±95% CI)	$\hat{\beta}$ (±95% CI)	R^2	$\hat{\alpha}$ (±95% CI)	$\hat{\beta}$ (±95% CI)	R^2
S03	3.6 (±0.6)	0.6 (±0.2)	0.99	3.0 (±0.8)	0.7 (±0.4)	0.98
S08	4.6 (±0.5)	0.6 (±0.2)	0.99	3.1 (±1.3)	0.6 (±0.5)	0.94
S11	4.1 (±0.5)	0.4 (±0.1)	0.99	2.5 (±0.2)	0.3 (±0.0)	1.00
S12	2.9 (±1.4)	0.5 (±0.3)	0.94	1.0 (±0.6)	0.6 (±0.2)	0.99
S13	4.1 (±1.0)	0.6 (±0.3)	0.97	2.0 (±1.2)	0.4 (±0.2)	0.95
S14	4.4 (±0.9)	0.4 (±0.1)	0.97	3.3 (±1.2)	0.4 (±0.2)	0.95
S15	3.2 (±1.0)	0.5 (±0.2)	0.97	1.4 (±1.1)	0.6 (±0.3)	0.95
S16	2.2 (±0.5)	0.9 (±0.3)	0.99	2.4 (±1.2)	0.5 (±0.3)	0.95
S17	3.2 (±0.8)	0.4 (±0.1)	0.98	0.7 (±0.3)	0.4 (±0.1)	1.00
S18	3.4 (±1.1)	0.5 (±0.3)	0.95	2.6 (±0.8)	0.4 (±0.1)	0.98
S19	2.5 (±0.6)	0.9 (±0.4)	0.98	1.7 (±0.6)	0.8 (±0.3)	0.99
Mean (SD)	3.5 (0.8)	0.6 (0.2)	0.97	2.2 (0.9)	0.5 (0.2)	0.97
Median	3.4	0.5	0.98	2.4	0.5	0.98

Curriculum Vitae

Name:	Renita Sudirga
Post-secondary Education and Degrees:	The University of Western Ontario London, Ontario, Canada 2009-2014 Ph.D. Queen's University Kingston, Ontario, Canada 2004-2009 M.Sc. Eng Queen's University Kingston, Ontario, Canada 2000-2004 B.Sc. Eng
Related Work Experience	Teaching Assistant The University of Western Ontario 2009-2013 Research Assistant The University of Western Ontario 2009-2014

Presentations:

Sudirga, R., Cheesman, M. and Macpherson, E. ”Effects of room-acoustic exposure on localization and speech perception in cocktail-party listening situations”, *J Acoust Soc Am,* 2014;135:2366 (Abstract) Presented in: 167th Meeting of the Acoustical Society of America, Providence, Rhode Island, May 5-9, 2014

Sudirga, R., Taylor, R. , Macpherson, E. and Cheesman, M. "Speech Communication in Vehicles: Intelligibility and Listening Effort". Poster presented at: Canadian Association of Speech-Language Pathologists and Audiologists Conference, Montreal, April 27-30, 2011

Sudirga, R., Brathwaite, M., Parsa, V., Macpherson, E. and Cheesman, M., “ Electroacoustic measurements of hearing aids in automobile listening environments”, Poster presented at: AUTO21 Annual Conference, Toronto, May 22-23, 2013