Effect of dynamic range compression on attending to sounds based on spatial location

by

Andrew H. Schwartz

B.S. Computer Systems Engineering
Boston University, 2005
S.M. Electrical Engineering and Computer Science
Massachusetts Institute of Technology, 2010

SUBMITTED TO THE DEPARTMENT OF HEALTH SCIENCES AND TECHNOLOGY IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY IN SPEECH AND HEARING, BIOSCIENCE AND TECHNOLOGY
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

SEPTEMBER 2013

Signature of Author:____
Department of Health Sciences and Technology
August 14, 2013

Certified by:____
Barbara Shinn Cunningham, PhD
Professor, Biomedical Engineering, Boston University
Thesis Supervisor

Accepted by:____
Emery Brown, MD, PhD
PhD/Director, Harvard-MIT Program in Health Sciences and Technology/Professor of Computational
Neuroscience and Health Sciences and Technology

Effect of dynamic range compression on attending to sounds based on spatial location

by

Andrew H. Schwartz

Submitted to the Department of Health Sciences and Technology on August 14, 2013 in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Speech and Hearing, Bioscience and Technology

ABSTRACT

Many hearing aids introduce nonlinear compressive gain to accommodate the reduced dynamic range that often accompanies hearing loss. Unfortunately, when applied independently at either ear, this gain can introduce fluctuations in interaural level difference (ILD), which is an important cue for spatial perception and attending to sounds in an acoustic mixture. Moreover, natural sounds produce complicated interactions between different sounds in a mixture, as a compressor's gain is driven by whichever source dominates the mixture within a specified temporal window. While independent compression can interfere with spatial perception of sound, it does not always interfere with localization accuracy or speech identification.

This thesis investigates the role of dynamic range compression on the ability to attend to target speech in the presence of interfering speech. First, the fundamental concepts behind dynamic range compression and its use are introduced, and used to develop a framework to understand some of the possible effects on ILD and spatial perception. This framework is applied toward the interpretation of the existing literature regarding dynamic range compression and spatial perception, bringing together a seemingly contradictory range of results. In particular, the framework presented here predicts that dynamic range compression will only affect performance in tasks for which relatively small spatial separations are tested, whereas many existing studies compare only large spatial separations to no spatial separation.

We describe and analyze the results of an experiment designed to test this prediction by systematically varying the spatial separation between different speech sources that normal-hearing listeners attended to. We found a robust but modest detrimental effect of dynamic range compression on listeners' performance. Linking the left and right compressors so that ILD was unaltered restored performance. Lastly, we develop a model to describe the utility of ILD for such tasks. The results of this model provide insight into the reported behavioral results, and generate predictions for how hearing impairment may alter the observed pattern of results.

Thesis Supervisor: Barbara Shinn-Cunningham
Title: Professor, Biomedical Engineering, Boston University

I. Table of Contents

II. List of Figures

III. List of Abbreviations

CRM: Coordinate response measure
HRTF: Head-related transfer function
ILD: Interaural level difference
ITD: Interaural time difference
JND: Just-noticeable difference
RAU: rationalize arcsine units
ROC: receiver operating characteristic
SPL: sound pressure level
SNR: signal-to-noise ratio

Chapter 1

Introduction

1.1 Auditory scene analysis

In many everyday settings, such as outdoors near a busy street or inside a crowded room, listeners must parse a complex mixture of sound sources into auditory "streams" in order to direct attention to a desired sound source. How the human auditory system accomplishes this remarkable feat has been the subject of ongoing investigation for several decades (Cherry, 1953; Bregman, 1990; Bronkhorst, 2000). To many of us, this task can seem somewhat trivial: for example, at a bar or restaurant, we often focus in on the conversation at our own table with ease, effortlessly ignoring conversations at other tables or even our own table, as well as background music and the cheerful clinking of glasses and silverware. Yet looking at the sheer mathematics of this problem, one cannot help but develop a sense of awe for the sophistication of our auditory system.

Bregman (pp 5-6, 1990) nicely illustrates the difficulty of the problem with the following *gedankenexperiment*. Consider a lake, and at the edge of this lake you have dug two small channels. In each of these channels you place a small handkerchief, to be moved by the waves entering the channels. Now, looking only at the motion of these cloths, you aim to determine all that is occurring on the surface of the lake. How many boats are on the water? How large are they, and how far away? Is the wind blowing? Are the geese on the lake surface? Where are they?

While this task may seem quite impossible, it is in principle exactly the problem our auditory system faces, and solves, on a daily basis. Unfortunately, individuals with hearing impairment often have greater difficulty with such situations. In this thesis, we focus on one aspect of how our auditory systems accomplish this task. Specifically, we look into the role of spatial separation of sound sources in listeners' ability to direct attention to a desired sound source. We then investigate how a signal processing strategy known as "dynamic range compression" (DRC) may affect this ability.

In this chapter, we will build up an understanding of how the normal-functioning auditory system uses spatial cues to focus attention on a desired source, and describe how hearing loss affects this ability. In Chapter 2, we discuss some of the current signal processing strategies used by modern hearing aids, and illustrate ways in which these strategies may in fact be detrimental in some situations. We will more thoroughly review the existing literature regarding the effects of DRC on several aspects of spatial hearing in Chapter 3, using our framework from Chapter 2 to interpret a seemingly divergent range of results. We then test these ideas using behavioral experiments in Chapter 4, and model these results to produce useful predictions for a broader population in Chapter 5. Finally, we will summarize the contributions of this work in Chapter 6, discussing its insights and limitations, and outline how future research could build upon this work.

1.2 Spatial attention is useful

Spatial cues such as interaural level differences and interaural time difference (ILD and ITD, respectively) arise due to the different paths taken by a sound to reach either ear. These cues allow listeners to reliably detect changes in a sound's location to angles within as small as 1 degree in the azimuthal plane (Mills, 1958). For complex sounds such as mixtures of speech, spatial separation between the target and distracting sources can provide large benefits to a listener's ability to detect and/or attend to the desired target sound source (Shinn-Cunningham, 2005). There are several mechanisms that allow for this benefit.

Acoustic head shadow, that is, the attenuation in level of a sound at the far side of the head due to occlusion by the head, results in a "better ear" effect that can improve the audibility of the target. This effect changes the signal-to-noise ratio at either ear, resulting in an improvement that can be as large as 20 dB in some specific stimulus configurations (Zurek, 1993; Shinn-Cunningham et al., 2001). The neurally represented SNR is likely further enhanced by binaural processing that can take advantage of ILDs and ITDs in the acoustic mixture (Durlach, 1963; Akeroyd, 2004; Culling, 2007). Some of the largest benefits of spatial separation on speech understanding, however, have been demonstrated when the target speech is clearly audible and intelligible but is difficult to distinguish from competing sources, making it difficult to focus attention on the desired source (Brungart, 2001; Kidd et al., 2005a; Ihlefeld and Shinn-Cunningham, 2008). In these cases, the spectrotemporal structure of real-world sources such as speech (including periodicity, spectrotemporal continuity, and common amplitude modulation) first enables listeners to segregate sources from within an acoustic mixture (Carlyon, 2004). Listeners can then focus attention on the desired source based on higher-level perceptual features such as perceived location (Shinn-Cunningham and Best, 2008).

1.3 Hearing-impaired listeners suffer from deficits in spatial hearing

Hearing aid wearers often report difficulty following conversations in crowded or noisy environments (Robillard and Gillain, 1996; Gatehouse and Noble, 2004; Bridges et al., 2012). While this deficit is related to many issues that occur with hearing loss, such as reduced frequency selectivity and sensitivity to temporal fine structure of sound waveforms (Hopkins and Moore, 2011), one relevant factor is that in such situations, hearing-impaired listeners often show a reduced ability to take advantage of spatial separation compared to normal-hearing listeners (Bronkhorst and Plomp, 1989; Noble et al., 1995; Best et al., 2011). Providing better amplification to high-frequencies, where ILD is more prominent, has been shown to improve the benefit of spatial separation in some hearing-impaired listeners (Ahlstrom et al., 2009; Moore et al., 2010).

Unfortunately, as we will see in the following section, current hearing aids often corrupt ILD, and this effect may limit hearing-impaired listeners' ability to follow speech in crowded or noisy environments. Methods to enhance spatial cues have been shown to improve localization performance and speech intelligibility in stationary noise (Francart et al., 2011; Wiggins and Seeber, 2013). Relatively little work has been done, however, on the effects of such signal processing strategies for attending to speech in the presence of competing speech, where spatial cues often provide the most benefit. In such situations, hearing-impaired listeners show a wide range of abilities, with some subjects showing as good a benefit from spatial separation as normal hearing counterparts to some subjects showing no benefit of spatial separation whatsoever (Marrone et al., 2008a). Providing a more robust set of spatial cues to hearing impaired listeners may help these listeners make the most out of a given acoustic environment.

1.4 Dynamic Range Compression in hearing aids affects spatial hearing

Hearing aids routinely employ a signal processing strategy known as "dynamic range compression" (DRC). DRC operates by providing greater amplification to signals that have low intensity than to signals that have higher intensity. For example, a quiet whisper received at an average sound level of 40 dB SPL may be amplified by 20 dB to an average level of 60 dB SPL so that it is audible, while a loud shout received at an average level of 80 dB SPL may not be amplified at all. DRC is provided in hearing aids to alleviate the problem of limited dynamic range available to hearing-impaired listeners, which is often believed to result from the deterioration of an active, compressive gain mechanism in the inner ear (Moore, 2007). In some settings, hearing aids that provide DRC can improve speech intelligibility compared to aids that provide only linear amplification (Moore, 1996; Jenstad et al., 1999).

While the benefit of bilateral vs. unilateral hearing aids has been known for decades (Markides, 1982; Noble and Gatehouse, 2006), it is common for the left and right hearing aids to operate independently of one another. Unfortunately, when DRC is applied independently to both ears, it can alter the ILD present in the acoustic signal. When a sound has some non-zero ILD, it is relatively more intense in one ear vs. the other. Due to the nature of DRC, the less-intense signal will be amplified relatively more than the more-intense signal, leading to a reduction in ILD. For several reasons, it is not immediately clear how compression will affect the spatial perception of sounds. Human listeners have a remarkable ability to adapt to statically altered interaural cues (Bauer, 1966; Simon and Aleksandrovsky, 1997; Shinn-Cunningham et al., 1998), and therefore the changes in ILD caused by DRC may not be perceptually meaningful to a listener after some period of acclimatization. However, DRC in hearing aids does not uniformly reduce the ILD even within a single acoustic stimulus, as we will see in detail in the following chapter. Complicating this picture further is the fact that DRC schemes in hearing aids can take anywhere between a few milliseconds and several seconds to respond to a new stimulus level (see: Moore, 2008). As a result, the effects of DRC on ILD are not as simple as a static reduction. Instead, the change to ILD will be temporally variable and dependent on the timing between different sound sources in a mixture. We will explore the effect of dynamic fluctuations to ILD in more depth in later chapters.

Among the existing studies that examine the effect of DRC on spatial aspects of hearing, an interesting mix of results can be found. With DRC, both normal-hearing and hearing-impaired listeners required larger differences in ILD between two otherwise identical stimuli to be able to distinguish them apart from one another (Musa-Shufani et al., 2006). Subjectively, DRC can also lead to effects such as more broad or diffuse sound images, illusory movement, and split sound images (Wiggins and Seeber, 2011, 2012). However, studies often find no effect of DRC on localization accuracy (Keidser et al., 2006; Musa-Shufani et al., 2006). The effect of DRC on the spatial benefit of spatial separation of speech from stationary noise is unclear (c.f. Moore et al., 1992; Wiggins and Seeber, 2013), and a similarly confusing mix of results can be found for speech with interfering speech (c.f. Kalluri and Edwards, 2007; Marrone et al., 2008a).

In this thesis, we attempt to reconcile this variety of results found in the literature by looking in detail at how DRC can affect acoustic spatial cues, and how such effects may influence spatial perception in a variety of situations. We will build a framework that allows us to interpret these existing results in a consistent manner. Here we are primarily interested in the influence of DRC on a listener's ability to attend to speech in the presence of interfering speech, a situation which repeatedly emerges as one of the chief complaints of hearing-aid users (Kochkin, 2010, 2011). We will therefore design an experiment aimed at testing this framework in a manner relevant to this common type of acoustic setting. As a first study of these effects, we will only perform this experiment with normal-hearing listeners. Therefore, we will develop a binaural auditory model from existing components to replicate the trends observed in our experiments and to generate predictions for hearing-impaired listeners using typical DRC in hearing aids.

Chapter 2

Decision Theory Framework

2.1 ILD as a decision variable

In this chapter we will build a more detailed understanding of how DRC affects binaural acoustic signals. We will describe a simple but common DRC strategy, and describe its effects on the ILD of sounds. In this thesis, we are focused on the utility of ILD on selective attention tasks, such as attending to one voice in a crowd, rather than on localization *per se*. While these two tasks are intuitively related, good performance in either of these tasks does not imply good performance in the other (Gallun et al., 2008; Schwartz et al., 2012). Therefore it will be important to keep the distinction between localization and spatial attention in mind as we develop our understanding of this complex issue, and as we interpret the results presented in Chapters 3-5. This distinction will be further illustrated as we develop examples in this chapter.

The view we take is that the auditory system uses "bottom-up" grouping cues to extract perceptual sound "objects" from an acoustic mixture; it then uses higher level features such as perceived spatial location to direct attention toward relevant sound objects (Shinn-Cunningham and Best, 2008). We are primarily interested in the latter process; that is, the utility of spatial cues such as ILD to determine if a specific token of sound belongs to the target of attention or not. With this view, spatial cues need only bear differences sufficient to tell two or more sound sources apart; they need not necessarily accurately identify the sources' physical locations. While multiple cues, including non-spatial cues, can generally contribute to this distinction, here we focus simply on the utility of the ILD cue by itself. The role of ILD may be somewhat redundant in certain situations where other cues are sufficient to distinguish the sources in a mixture (e.g., Helfer and Freyman, 2008; Marrone et al., 2008b), but the presence and strength of any of these cues can depend largely on the stimuli and the environment. It is therefore useful to understand the role of ILD specifically, and how DRC may affect that role.

To address this question, we will treat ILD as a decision variable used to estimate if the target sound is present or not at any given instant. This type of problem is well-studied in decision theory, allowing us to use tools from this discipline to quantify the effect of DRC. While we will provide much more concrete data to this effect in our auditory model in Chapter 5, it suffices for now to describe this essential problem in more general terms as follows. We will describe the ILD associated with a specific frequency band of the target source and one or more interfering sources by a distribution of values, as illustrated in Figure 2.1, computed over any time window in which the sources are assumed to not be moving. This distribution can represent either a parametric probability distribution modeled from observations or a non-parametric summary of those observations. Given this set up, we can define a decision rule by partitioning the space of observable ILD values into two or more regions, each of which is associated with either the target or the masking sources. When a sound element's ILD is determined to lie in the region associated with the target, the decision rule declares that this sound object containing this element belongs to the target source. Note that other features, such as vocal identity and even ILD in other frequency bands, could provide additional or even conflicting information; in general, no single feature is a perfect predictor of a sounds source's identity. It is beyond the scope of this thesis, however, to consider the combination of such conflicting pieces of information. Instead, we focus solely on how good of a predictor ILD is, while noting that DRC may affect other cues in different ways.

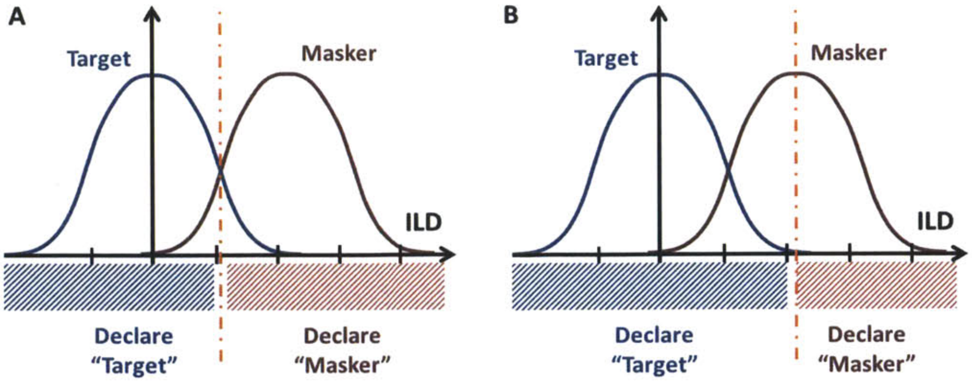

Figure 2.1 – Illustration of two of infinitely many possible decision rules given identical ILD distributions of the target (blue) and masker (red). These two rules illustrate the trade-off between two different types of error, dubbed "false alarm" and "miss" errors. A: This rule equally weights both error types: the probability of incorrectly declaring "target" equals the probability of incorrectly declaring "masker". B: This rule has almost no probability of incorrectly declaring "masker", as the masker decision region (red) contains virtually no probability mass from the target ILD distribution. However, this rule has a rather high probability of incorrectly declaring "target", as the target decision region contains a significant portion of the masker probability distribution mass.

We will apply more formal techniques to this type of formulation in Chapter 5; for now, however, it suffices to say that the utility of ILD to distinguish between the target and masker is inversely related to how much these two distributions overlap. If there is no overlap, then the decision regions for the target and masker can cover all areas where the corresponding distribution has non-zero probability, and ILD can be used to uniquely determine which source generated a given sound. This is not realistic, and in general these distributions will overlap by some amount. This overlap can be due to acoustic factors, such as reverberation and the frequency-dependence of ILD, as well as neural factors limiting the precision of any neural representation of these distributions. As the overlap increases, we can say that ILD is becoming less useful for determining which source a particular sound belongs to. When there is full overlap (i.e., when the target and masker distributions are identical), then using ILD to distinguish between the two sources is no better than guessing at random.

Notice that the overlap between two distributions is determined by the relative location of the peaks as well as the widths of the distributions, and is independent of the absolute position of the center of either distribution given these two measurements. Thus this framework is ideal for describing the discriminability between two stimuli rather than the location of either. Localizing these stimuli would instead be accomplished by, for example, estimating the mean of either distribution and mapping the resulting ILD value to a corresponding physical location. We will however not directly address the question of localization in this thesis, focusing instead on the use of spatial cues to direct attention toward a target sound source.

2.2 DRC: Steady-state response

Now that we have defined our problem setup, we can begin to assess the effects of DRC on the discriminability of sounds by ILD. There are many varieties of DRC algorithms used in hearing aids and other audio applications, but they all largely share a common feature: a simple compression scheme will apply linear gain below some set threshold, and will provide progressively less gain as the input signal increases in level beyond this threshold. Such a compression scheme is illustrated in Figure 2.2. The amount of gain applied with increasing input levels is determined by the compression ratio R; the compressor aims to increase the output signal level by 1 dB for every R dB increase in the input signal level. Thresholds on compressors designed for speech intelligibility are often set to be around the level of relatively quiet speech, but can vary from user to user. Typical values of R for hearing aids range from just above 1 (almost linear) to roughly 3 for more severe cases of hearing loss. Commonly added features to compression schemes include a sharp roll-off for very low levels so as not to over-amplify ambient background noise and microphone noise, a return to linear gain for high levels to avoid providing negative gain to high-level sounds, and a hard-limiter for very high levels to prevent dangerously loud sounds from entering the listener's ear. However, for this thesis, we will focus on the simple compression scheme defined by a linear gain, compression threshold, and compression ratio (as well as temporal dynamics; see Sec.2.3).

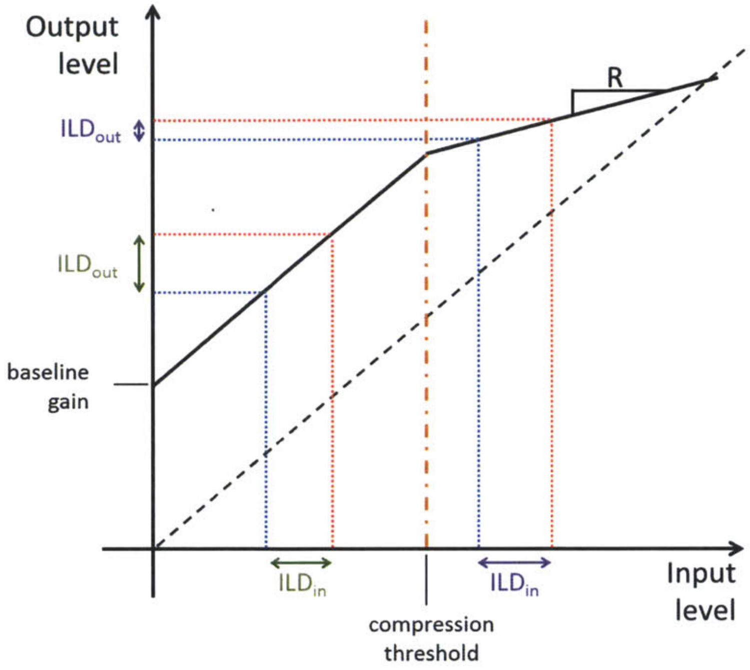

Figure 2.2 – Input-output plot of a simple compression scheme, represented by the solid black line, for arbitrary decibel values. The dot-dash line shows the compression threshold. For input levels below this threshold, sounds are linearly amplified: constant gain in decibels is applied. Above this threshold, a compression ratio of R is applied: every increase of R dB in the input results in a 1 dB increase in the output level. Red and blue dotted lines represent the right and left inputs, respectively, of two example stimuli at different overall levels. These two examples have the same input ILD (represented by distance between the red and blue lines on the x-axis, labeled "ILD$_{in}$"), but only the stimulus whose level is above the compression threshold has a reduced output ILD (distance between the red and blue lines along the y-axis, , labeled "ILD$_{out}$").

As illustrated by Figure 2.2, when a binaural sound is compressed by this type of processing operating independently on both left and right-ear signals, its ILD can be reduced by up to a factor of R (the compression ratio). Thus the steady-state response of bilaterally-independent DRC is to reduce the ILD of most sounds in a listener's environment, bringing sounds closer together along this perceptually relevant dimension. There are at least two reasons why the overall reduction of ILD might greatly not affect the perception of auditory space in hearing-aid wearers. The first is that listeners are rather good at adapting to modified interaural cues; such adaptation can be shown in behavioral responses (Bauer, 1966; Shinn-Cunningham et al., 1998) as well as on a neural level (Kacelnik et al., 2006; Dahmen et al., 2010). If hearing aids were to universally reduce ILD, listeners wearing those aids would likely to be able to learn the mapping of the altered ILDs to physical locations. However, this reduction in ILD occurs only for sounds above the compression threshold, and it is unknown how well listeners could adapt to a level-dependent change in ILD.

Even if listeners are able to adapt to modified ILD, we should still expect a reduced ability to discriminate between two ILD values that are very close to each other. Going back to our decision theory setup of Figure 2.1, let us assume for a moment that some part of the width of the ILD distribution is due to internal noise or precision limits of the underlying neural system (e.g., Durlach and Braida, 1969; Searle et al., 1976) and that this component of neural noise cannot be altered by external stimulus processing. Now imagine two sounds whose respective ILDs are just barely distinguishable due to this noise. When those ILDs are reduced by DRC, then those two sounds will cease to be distinguishable by ILD. This fundamental argument applies regardless of an individual's ability to adapt to a new mapping of ILDs to physical locations in space.

Another argument for why this reduction might not affect spatial perception is specific to hearing-impaired listeners: if the mechanism of hearing loss reduces compressive gain in the inner ear, ILD will be expanded, and hearing aid DRC should approximately restore "normal" ILD values. Due to the varied and complicated nature of hearing loss and the practicalities of hearing aid devices, this restoration can only be approximate; nevertheless, on a more practical note we need simply to point out that the same argument as stated in the preceding paragraph applies. That is, for ILDs that are just distinguishable from neural noise, a reduction of those ILDs will render them indistinguishable regardless of whether or not that reduction can be said to be restoring normal neural ILDs. Moreover, despite any expansion of ILD due to hearing impairment, hearing-impaired listeners generally have comparable or worse ILD sensitivity compared to normal-hearing listeners (Florentine et al., 1993; Koehnke et al., 1995; Musa-Shufani et al., 2006), and compression indeed degrades this sensitivity (Musa-Shufani et al., 2006).

A further complication to our increasingly complicated picture is that DRC will not uniformly reduce ILDs in any stimulus or mixture of stimuli. Even when a target sound source is presented alone, fluctuations in its level interacting with a given DRC scheme can cause fluctuations in ILD, and these fluctuations are likely the cause of perceived movement, diffuseness, and splitting of ordinarily stationary sounds for normal-hearing listeners (Wiggins and Seeber, 2011, 2012). When interfering sources are also present, the mixture of both stimuli can cause the compressors to affect ILD differently than when the target is presented alone. Here, the fluctuations to ILD can become unpredictable and not directly related to the target. Therefore, experiments that investigate the effect of DRC on isolated stimuli (e.g., Keidser et al., 2006; Musa-Shufani and Walger, 2006; Wiggins and Seeber, 2012), or even stimuli with stationary noise (e.g., Moore et al., 1992; Wiggins and Seeber, 2013) may underestimate the effect of DRC compared to what may be revealed when multiple, dynamic stimuli are present, such as in a speech mixture.

2.3 DRC: Temporal dynamics

Fluctuations in ILD due to fluctuations in a stimulus' level and in the level of interfering stimuli are still only a part of the story. Another factor of hearing aid DRC that may dramatically influence its effect on spatial perception comes from its temporally-varying response. As mentioned above, naturally occurring sounds, such as speech, fluctuate in level from moment to moment. These fluctuations are important to speech intelligibility (Elhilali et al., 2003). If DRC were to act instantaneously or nearly instantaneously, these modulations could easily become severely corrupted and speech intelligibility can suffer as a result (Stone and Moore, 2004). Furthermore, instantaneous compression would distort the fine-structure of an acoustic waveform, introducing significant harmonic distortion. To avoid these detrimental effects, hearing aids smooth their responses to fluctuating input levels using "attack" and "release" time constants (ANSI, 2003) on the order of milliseconds to seconds. The attack and release time constants determine how quickly the gain of the compressors changes in response to a sudden rise or drop, respectively, in the input signal level.

These temporal dynamics add another dimension of complexity to the nature of DRC-imposed ILD fluctuations. Some examples of this are illustrated in Figure 2.3, using an attack and release time constant of 40 ms. In panel A, the ILD at the onset of an isolated tone burst is unaffected by DRC. Because localization is dominated by spatial cues near stimulus onsets (Freyman et al., 1997; Stecker and Hafter, 2002), localization may be relatively unaffected by compression in such simple conditions, consistent with the previous reports of a lack of effect of compression on localization of isolated stimuli in quiet, anechoic environments (Keidser et al., 2006; Musa-Shufani et al., 2006).

In panel B, a tone burst from center follows a tone burst from the right. As in panel A, the source from the right has a "clean" onset ILD. The source from the center, however, has an ILD at its onset that is pushed farther left (negative ILD) when DRC is applied compared to when it is not. At the onset of this second stimulus, the compressors are still applying gain according to the input levels of the first stimulus; that is, they are applying a negative bias to the ILD. Therefore, for some time following the onset of the second stimulus, the ILD at the output of the compressors will be negatively biased regardless of the ILD at their input. While this effect may impair localization of the second stimulus, we note that the onset ILDs are actually farther apart with DRC than without. Thus, if we assume that discriminating between two source locations is in fact heavily influenced by stimulus onsets, as localization of single sources is, then it is plausible that DRC in this situation can actually enhance the perceptual distinction between these two stimuli, actually improving a listener's ability to discriminate between them. This situation again highlights the difference between DRC's possible effects on discriminating between different sound sources versus localizing them. Notice that the linear ILD trace also takes slightly longer to adjust to the new value; this is because the more intense stimulus continues to influence the power estimate used to compute ILD for a short period of time, which is computed over a 20 ms window (for more details on how ILD is computed here, see Section 5.1.2).

Finally, in panel C, we add more source locations: a source from the right is followed by a source from center, in turn followed by a source from the left, and finally the source from center again. The blue line, representing the uncompressed ILD, can be followed to determine which source is active at any given time on the plot. Here we can see that the ILD at the onset of the source from center (where the blue line goes to zero) is pushed in either direction by DRC (pink line), depending on the stimulus that preceded it. Even in this simple example it becomes difficult to use the compressed ILD plot alone to distinguish between the various sources locations. In more realistic settings, with any number of potentially overlapping stimuli, each with varying levels and more varied timing, the effects on ILD can become even more chaotic.

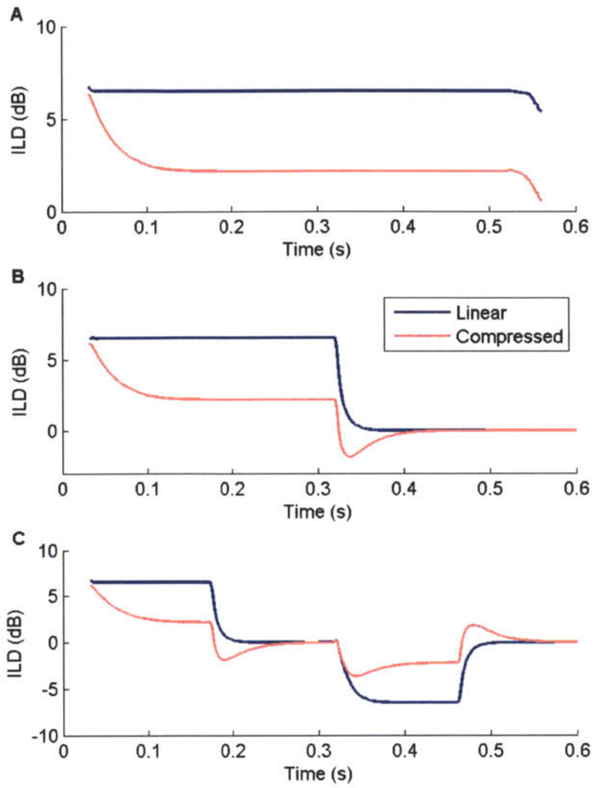

Figure 2.3 – ILD of simple example stimuli, under linear processing and bilaterally-independent dynamic range compression. In each panel, the stimulus begins after a 10 ms silence to avoid edge effects in the compressor's smoothing operation. A: Isolated tone burst at 70°. B: Tone burst from 70° followed immediately by a tone burst from center. B: Tone burst from 70°, center, -70°, then center again. ILD was computed using power estimates in a 10-ms exponential window, resulting in brief dynamic transitions even in the linear processing case.

Generalizing from these observations, we say that the bias in ILD applied to one stimulus will be applied for some time to the onset of a stimulus that immediately follows it. Lest the reader suspect that the story ends here, we further point out that how long the compressors take to adjust to the new stimulus depends on an interaction between the compression attack and release time constants and the level of both stimuli. This dependence is illustrated in Figure 2.4. In both panels, a stimulus from the right is followed by a stimulus from the center, and the compressor is set to have an attack time of 10 ms and a release time of 100 ms. The only difference between these two panels are the levels of each of the two stimuli. In panel A, the right stimulus is set to 50 dB SPL[1], while the center stimulus is set to 70 dB SPL. Because the level of the center stimulus is greater in both left and right ears than the level of the initial stimulus, the attack time is applied, and the ILD adjusts quickly to the steady-state value of the second stimulus' compressed ILD. In panel B, the levels of the two stimuli are reversed. Here, because the level of the second stimulus is now lower in both left and right ears, the longer release time applies, and the ILD takes longer to adjust.

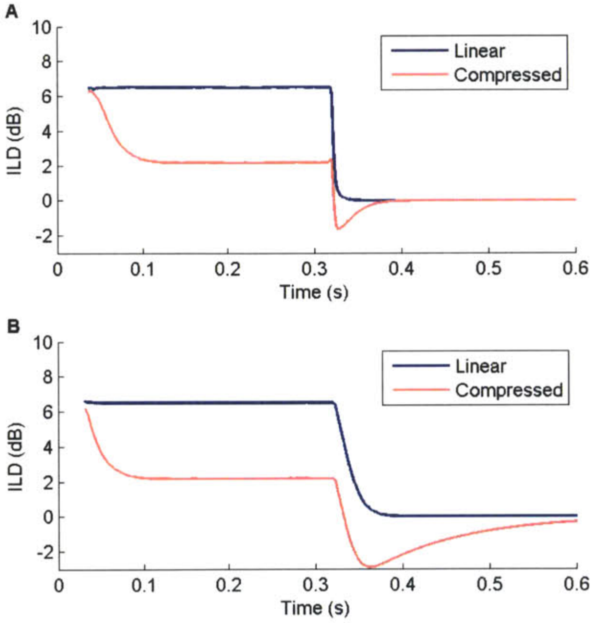

Figure 2.4 – ILD traces of example stimuli using a longer release time constant (100 ms) than attack time constant (10 ms). In panel A, the initial right stimulus has a level of 50 dB SPL, and the second center stimulus has a level of 70 dB SPL. In panel B, these levels are reversed. The profound difference in the pink trace, representing the compressed ILD, even 100 ms after the transition, is due to the attack time applying to the transition in panel A but the release time applying in panel B due to the differences in stimulus level.

---

[1] Sound level is set prior to spatial filtering. The spatial filters were calibrated relative to the "diffuse-field", which

2.4 Conceptual model of the effect of DRC on ILD discriminability

With the examples presented in the previous section it is becomes evident that the exact effects of DRC on ILD can quickly become quite complicated and mathematically intractable. Therefore, as a conceptual simplification to allow some principled analysis and exploration of the effects of DRC on the ILD of mixtures of multiple, dynamic stimuli, we consider the following big-picture effects of DRC on ILD: (1) overall, steady-state ILD reduction, and (2) ILD fluctuation, caused by the dynamics both within stimuli as well as in the compression parameters. In particular, as we saw illustrated in Figure 2.3, the fluctuations in ILD due to DRC can bias ILD in any direction, even if the uncompressed ILD is zero!

Building on this construct, we develop a simple, conceptual model of how we expect DRC to affect the utility of ILD in such tasks as attending to target speech in a mixture. Returning to our decision theory setup, we assume we have an ILD distribution associated with the target stimulus and another associated with one or more masking stimuli. The mean location of these distributions will be determined by the physical locations of the corresponding stimuli. The width of these distributions will depend on internal noise related to the representation of ILD as well as fluctuations in ILD over time. We are interested in modeling the effects of ILD reduction and ILD fluctuations caused by DRC.

ILD reduction will in general bring the mean of these distributions toward zero. It can also reduce ILD fluctuations present in the acoustic stimulus itself. However, reduction of ILD might not reduce internal noise, and therefore the net result of moving the ILD distributions closer to zero can be an increase in their overlap. As we saw in the previous section, ILD fluctuations can occur in any direction regardless of the unaltered source ILD, so we describe these additional ILD fluctuations by an increase in the distribution width. This effect will further increase the overlap between target and non-target ILD distributions. These possible effects are illustrated in Figure 2.5.

Notice that the ILD distributions of the blue target at center and the red masker from 90° have practically no overlap with or without the effects of DRC. Thus, in this case, the conceptual model predicts no effect of DRC on the discriminability of these two sources: the distributions are sufficiently distinct such that they are robust to the modest degradation illustrated here. The green distribution for the 15° masker, however, overlaps that of the target substantially more with DRC than without; the conceptual model therefore predicts that for stimuli that are "sufficiently close", as these stimuli are, DRC will result in impaired discriminability.

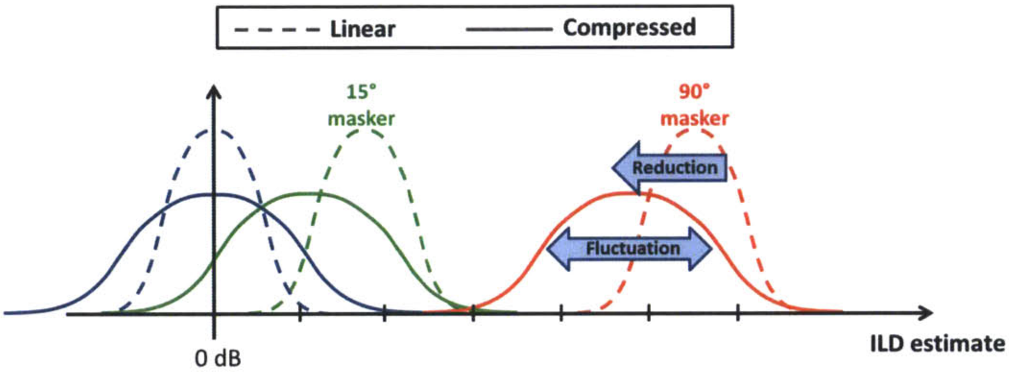

Figure 2.5 – Cartoon illustration of possible effects of dynamic range compression on ILD distributions for three sources. Dashed lines represent ILD distributions prior to DRC, solid lines represent distributions after DRC is applied. (1) Blue: a target source from center, with mean 0 ILD. (2) Red: a masker from far right (90°). (3) Green: a masker from an intermediate location (15°). Arrows mark the two groups of effects: ILD reduction and ILD fluctuations. Reduction brings distribution means closer to zero, while fluctuations result in wider distributions (even for sources with zero mean ILD).

Figure 2.5 illustrates that bilaterally independent DRC may impair the ability to attend to a source in an acoustic mixture in some, but not all, situations. Specifically, when sources are far enough apart in space such that the illustrated effects on ILD do not result in any notable increase in the overlap between two ILD distributions (e.g., blue and red distributions in Figure 2.5), then the ability to attend to either source based on ILD is unlikely to be affected by this corruption. This insight highlights an important design consideration for experiments that investigate the effect of DRC in real-world settings. Many studies quantify the benefit of spatial cues by comparing performance in some selective attention task when stimuli are collocated vs. separated by a large angle, such as 90° (e.g., Marrone et al., 2008a). However, our conceptual model shows that performance with large separations is unlikely to be affected by DRC, but that performance might be affected for smaller target-masker separations. Therefore, the choice of studying only large separations can lead to the erroneous conclusion that DRC has no impact on spatial listening in real-world situations, which can include smaller spatial separations between sound sources.

Chapter 3

Interpretation of existing literature

In this chapter, we present a thorough review of the existing literature that deals with the effects of DRC on spatial hearing. We show that this conceptual framework of Chapter 2 can explain the variety of results found in the literature despite their apparent conflicting interpretations.

3.1 Spatial attention

Of primary interest in this thesis is the effect of DRC on a listener's ability to attend to a target speaker in the presence of competing speech. Two studies address this issue directly, measuring the benefit of spatial separation in terms of a reduction in the SNR required to successfully attend to a target when maskers were spatially separated from the target compared to when they were collocated with the target. One of these studies showed evidence of a detrimental effect of DRC for normal-hearing listeners when ILD was the only available spatial cue (Kalluri and Edwards, 2007). There was little to no effect when ITD was also available. Oddly, hearing-impaired listeners in the study demonstrated no benefit of spatial separation at all, and were therefore not tested with compressive amplification. In the other study, there was no effect of compressive amplification for hearing-impaired listeners (Marrone et al., 2008a). A caveat to this latter study is that the comparison was made between aided and unaided (but equal sensation level) listening, so effects due to DRC cannot be decoupled from effects due to other device-related factors such as frequency shaping and microphone sensitivity. For instance, speech may have been more intelligible due to better frequency shaping with the hearing aid devices, counteracting a possibly detrimental effect of DRC on spatial hearing.

The results of both of these studies should be interpreted carefully. Both used speech material from the Coordinate Response Measure corpus (CRM; Bolia et al., 2000). This corpus contains sentence exclusively of the form "Ready [callsign] go to [color] [number] now." Unfortunately, as a result, when multiple sentences from this corpus are used simultaneously, the individual words will exhibit an unnaturally high degree of temporal overlap. Such a situation is known to limit the benefit of spatial cues, particularly for hearing-impaired listeners (Best et al., 2011). Therefore, any potential effect of DRC may have been masked by the inability of listeners to use spatial cues as effectively as might have been possible had less constrained, running speech been used. Consistent with this concern, hearing-impaired listeners in the study of Marrone et al. showed an average of only 3 dB benefit of spatial separation, and hearing-impaired listeners in the study of Kalluri and Edwards showed no benefit at all. By comparison, studies using either other speech corpora or signal processing aimed at reducing the spectrotemporal overlap between target and masker speech often show average benefits between 6 dB and 10 dB for hearing-impaired listeners (Arbogast et al., 2005; Neher et al., 2009, 2011).

Another concern with the high degree of temporal synchrony between sentences using the CRM corpus specifically involves the nature of bilaterally independent DRC and the fact that these studies placed maskers symmetrically flanking the target (i.e., with equal but opposite azimuths or ILDs). If two symmetrically placed maskers are active simultaneously in a particular frequency band, then the level at the input of the left and right compressors will be approximately equal, and therefore there will be no effect on ILD. Due to differences in the words spoken by each talker, there may still be some differences in level at different frequencies, but it is nevertheless likely that this situation does not fully represent the range of real-world signals experienced by listeners. Kalluri and Edwards addressed this artificial symmetry by attenuating one of the maskers by 6 dB; still, the net effect here compared to equal-level maskers is just a constant bias in ILD. To truly represent real-world settings, the temporal envelope of the target and maskers should be roughly independent, which it likely to increase the effect of DRC on ILD.

A final concern with the study of Marrone et al. is that spatial benefit was only measured for maskers located at ±90°. As illustrated in Figure 2.5, this scenario may underestimate the practical consequence of DRC on spatial hearing relevant to everyday listening. While the limited benefit of spatial separation shown in this study may have necessitated this large separation, it is possible that using less constrained stimuli such as running speech would have allowed for a larger benefit even at smaller angles, which might reveal an effect of DRC.

3.2 ILD just-noticeable-differences

DRC can lead to a large increase (worsening) in just-noticeable-differences (JNDs) for ILD (Musa-Shufani et al., 2006) of narrow-band noise bursts. That is, DRC leads to an increase in the amount by which the ILD of two stimuli must differ in order for listeners to reliably tell them apart. This is perfectly consistent with our central argument: when two stimuli are "sufficiently close" in ILD, such as when they are just-noticeably different, then DRC will impair a listener's ability to discriminate between them based on ILD.

This effect, however, was limited to when compression employed fast time constants. This observation suggests that listeners were able to use the uncompressed ILD at the stimulus onsets to improve their performance when the time constant was long. Even when the attack time was as short as 2 ms, JNDs were increased by less than the compression ratio (either 3:1 or 8:1), implying that even a very short onset ILD can be useful. While Musa-Shufani and Walgner did not perform pair-wise tests for each attack time, the data suggest that there may still be an effect with longer attack times. If this effect is real, it implies that despite having access to the cleaner onset ILD, listeners were still unable to completely ignore the ongoing, reduced ILD.

3.3 Localization

Despite the effect on JND, in some situations, DRC has no effect on localization accuracy of narrow-band or wide-band noise bursts (Keidser et al., 2006; Musa-Shufani et al., 2006). One study did show that compressive amplification increased localization errors of various broadband stimuli, including noise bursts and a telephone ring (Van den Bogaert et al., 2006); but, as we noted with the speech masking experiment of Marrone et al., it was not DRC *per se* that was studied but compressive amplification vs. unaided performance.

There are several reasons why the first two studies may have shown no effect of DRC on localization accuracy. First, the spatial separations between loudspeaker locations used in these experiment, and therefore the minimum azimuthal difference between different stimuli, were 15° or 18° respectively, at least an order of magnitude greater than listeners' average JNDs for azimuthal separation. It is possible, then, that the stimuli presented were not "sufficiently close" to affect the correct identification of the sound source locations used; the effect of DRC might be so small as to only matter for the smallest of spatial separations or under special conditions (the study by Van den Bogaert et al. also used spatial separations of 15° and did find an effect). However, there are still some more considerations when interpreting these results.

One of the most important reasons why DRC might not have had a significant impact on localization performance in the first two studies is that these experiments used stationary stimuli in quiet environments without distracting sound sources. By contrast, the study performed by Van den Bogaert et al. included localizing a telephone ring in the presence of speech babble. Keidser et al. used one condition with stationary background noise, but unfortunately did not include a comparison of DRC with linear amplification in this condition. From our framework above, then, we have two reasons to suspect that this overly simple experimental condition may have failed to show an effect of DRC. The first is that the onset ILD will have been unaltered between linear and compressed conditions, and as we have seen with the JND experiments of Musa-Shufani et al., this onset ILD can play a crucial role in performance. The second is that fluctuations in the target level as well as the addition of interfering stimuli results in the compressors having more complicated effects on ILD. Neither of these effects was present in the stimuli used in these two studies.

3.4 Speech intelligibility in noise

One study objectively measured speech intelligibility by hearing-impaired listeners in pseudo-stationary noise (Moore et al., 1992). By pseudo-stationary noise, we refer to 24-talker babble; the benefit of spatial separation for speech on speech masking diminishes as the number of interfering talkers increases beyond 2, and this effect appears to be due to the fact that a mixture of interfering talkers becomes more similar to stationary noise as the number of independent sources increases (Freyman et al., 2004). In this case the action of the compressors due to the maskers will be nearly stationary as well, limiting its effect on the variations in the ILD of the target. This pseudo-stationarity was particularly relevant to the hearing aid used in this experiment because the compressors operated only on two independent frequency channels. Using two relatively wide channels makes it highly likely that a given channel will be receiving energy from several of the 24 independent speech utterances at any given time, resulting in relatively stationary behavior over time. Additionally, for the babble to be a comparable level to the target speech, the level of each individual talker in the babble had to be low so as to be not easily confused with the target, resulting in a target that was much louder than, and therefore perceptually very distinct from, each talker in the babble. This perceptual distinction reduces the need for cues such as ILD in order to distinguish between the target and masking speech (Brungart, 2001). Note, finally, that the 24-talker babble was separated into 12 talkers in the right speaker at +90°

and 12 talkers in the left speaker at -90°, which we have argued may not be an ideal setup to test the effect of DRC on spatial perception.

In a more recent study involving normal-hearing listeners, DRC was shown to have a detrimental effect on speech intelligibility in stationary noise that was separated from the target speech by 60° (Wiggins and Seeber, 2013). Linking the compressors to provide identical gain at any given time instant almost fully restored performance. This effect of independent DRC, however, was attributed not to spatial perception but simply to a decrease in the long-term, effective SNR when compression was bilaterally independent. More precisely, when stationary noise is used, the noise level remains roughly constant, and therefore epochs with higher SNR typically imply a higher overall level of the mixture. As a result, these high-SNR regions of the mixture will generally be attenuated more than low-SNR regions, reducing the overall SNR. Wiggins and Seeber also showed how linked compression restored the long-term SNR, due to the action of the compressor at the better ear being driven largely by the noise in the worse ear, resulting in the signal at the better ear having a long-term SNR almost equivalent to the linear condition.

When a speech masker is used, however, this effect may not persist as the noise level will not be constant. In this case, higher target-to-masker ratios do not necessarily correspond to higher SNR, and thus it is less clear how to apply the arguments above. Moreover, intelligibility is not usually a limiting factor in performance for such situations (Brungart, 2001), and therefore the modest gains to SNR that are discussed by Wiggins and Seeber is not likely to be a critical factor in performance. Listeners are able to take advantage of short-term glimpses of good SNR with speech-on-speech masking (Brungart and Iyer, 2012), but the SNR within these short-term glimpses is not altered by DRC (as it cannot amplify the target and masker differently at a given time instant).

3.5 Subjective qualities of spatial hearing

Normal-hearing subjects reliably report adverse effects of DRC on a set of subjective metrics of spatial hearing for a range of stimuli (Wiggins and Seeber, 2012), including speech as well as noise and tone bursts with gradual onsets. Specifically, these metrics were derived from a series of eight questions, reduced by hierarchical cluster analysis to four qualities: diffuseness, movement, image-split, and externalization. Reports of diffuseness, movements, and image-splits were all increased when independent DRC was used, particularly for stimuli that were high-pass filtered to emphasize the contribution of ILD over ITD. Stimuli with sharp onsets, such as a noise burst and pulse trains, were largely unaffected on these metrics. This difference is consistent with the idea that, in quiet conditions, listeners can make use of the clean spatial cues at stimulus onsets, and also consistent with the lack of effect of DRC on localization of such sounds (Keidser et al., 2006; Musa-Shufani et al., 2006).

These adverse effects observed with speech and slow-onset stimuli were absent when a static ILD bias, which gave the same overall magnitude of ILD reduction but without the temporal effects discussed in Section 2.3, was used in place of DRC. This result confirms the role of the temporal dynamics of compression in degrading spatial hearing, at least for normal-hearing listeners. Note also that this study did not include any conditions with background or interfering sounds, which may have increased these effects further.

Because the temporal dynamics of compression were critical to these effects, it is likely that hearing-impaired listeners would be susceptible to the same effects, despite the argument that DRC would restore mean ILDs to normal. However, there are other reasons why we might expect hearing-impaired listeners to be affected differently by DRC. For example, older hearing-impaired listeners can be insensitive to changes in source width (Whitmer et al., 2012, 2013). Source width seems intuitively related to diffuseness; however, in these experiments, source width was varied by adjusting the interaural coherence, which is related to timing cues in the stimulus waveform rather than ILD. It is possible, then, that these listeners might rely more heavily on the stability of ILD cues for precise spatial percepts, and therefore might suffer relatively more than normal-hearing listeners when ILD is corrupted. On the other hand, if the reported insensitivity to image width represents a similar perception of image diffuseness despite clean ILD, then corruption to ILD may not result in any additional deterioration of performance. Even if this latter idea proved true, however, younger hearing-impaired listeners showed better sensitivity to source width as conveyed by interaural coherence, and therefore may suffer from effects related to increased diffuseness due to DRC.

3.6 Predictions for listening experiments

The previous chapter explores some of the possible effects of DRC on ILD. The current chapter reviews a wide variety of published literature, interpreting the results under the general framework developed in Chapter 2; as a result, we are able to resolve many of these apparent discrepancies in past findings. In particular, we note two summarizing observations: (1) DRC should have the strongest effect for discrimination of small or moderate angles, and likely does not affect discrimination of large angles such as 90° used in many experiments; and (2) the adverse effects to spatial perception are likely to be worse in complex environments, such as those with dynamically fluctuating interfering noises. We therefore designed a set of experiments described in Chapter 4 that takes these observations into account in order to explore where the effect of DRC may be maximally adverse.

In our experiments, we investigated the effect of DRC on performance in a task where listeners attended to target speech in the presence of competing speech. The timing of the speech tokens was designed so that target and masker words were not overly synchronized as in some of the aforementioned studies. We measured the minimum amount of spatial separation required for threshold performance in this task, hypothesizing that bilaterally-independent DRC will increase this required amount of separation.

We also tested a condition wherein the left and right compressors are linked instead of operating independently, which may alleviate adverse effects on spatial perception by preserving ILD.

For our experiments, we used head-related transfer function (HRTFs) to preserve natural occurring spatial cues including ITD. One reason that some experiments fail to show an effect of DRC on spatial perception may be the presence of ITDs, which are generally unaffected by DRC. Although listeners cannot optimally combine corrupted and uncorrupted spatial cues (Ihlefeld and Shinn-Cunningham, 2011), the presence of uncorrupted ITD cues may provide sufficient robustness to spatial perception such that no overall effect is observed. In general, ITD will be available to listeners in real-world settings. However, ITDs in realistic settings can themselves be corrupted by interference with other sound sources as well as reverberation (Shinn-Cunningham et al., 2005; Monaghan et al., 2013). In such settings, the redundancy provided by having multiple perceptual cues to which listeners can direct attention is likely to allow for robust performance, and further degrading a cue such as ILD may interfere with performance or make listening more effortful in some situations. Therefore, an experiment designed to provide ILD as the only selection cue available to listeners (e.g., Kalluri et al., 2007) may be informative despite the lack of real-world cues such as ITD. Nevertheless, to better represent real-world settings typically encountered by listeners, we chose to keep realistic ITDs in our experiments.

Chapter 4

Psychophysical Experiments

This chapter describes psychophysical experiments designed to explore the effect of DRC on spatial perception when multiple sources that are spatially separated but spaced relatively close together (compared to 90° as in many previous experiments). These experiments have been documented and either published or submitted for publication; we therefore include the full text of these manuscripts here. In the first manuscript, we describe a series of experiments designed to explore such an effect with varying compression and room conditions. In the second, shorter manuscript, we describe a follow-up experiment where we used ITD as the only cue that separated the target from the maskers. This second manuscript provides an informative control case that clarifies the mechanisms responsible for the effects demonstrated in the first manuscript.

4.1 Effects of dynamic range compression on spatial selective auditory attention in normal-hearing listeners

Andrew H. Schwartz

Harvard / Massachusetts Institute of Technology, Speech and Hearing Bioscience and Technology Program, Cambridge, Massachusetts 02139

Barbara G. Shinn-Cunningham[2]

Center for Computational Neuroscience and Neural Technology and Biomedical Engineering, Boston University, Boston, Massachusetts 02215

Running title: Spatial attention with bilateral compression

---

[2] Author to whom correspondence should be addressed. Electronic mail: shinn@cns.bu.edu

4.1.1 Abstract

Many hearing aids introduce compressive gain to accommodate the reduced dynamic range that often accompanies hearing loss. However, natural sounds produce complicated temporal dynamics in hearing aid compression, as gain is driven by whichever source dominates at a given moment. Moreover, independent compression at the two ears can introduce fluctuations in interaural level differences (ILDs) important for spatial perception. While independent compression can interfere with spatial perception of sound, it does not always interfere with localization accuracy or speech identification. Here, normal-hearing listeners reported a target message played simultaneously with two spatially separated masker messages. We measured the amount of spatial separation required between the target and maskers for subjects to perform at threshold in this task. Fast, syllabic compression that was independent at the two ears increased the required spatial separation, but linking the compressors to provide identical gain to both ears (preserving ILDs) restored much of the deficit caused by fast, independent compression. Effects were less clear for slower compression. Percent-correct performance was lower with independent compression, but only for small spatial separations. These results may help explain differences in previous reports of the effect of compression on spatial perception of sound.

PACS codes: 43.66.Ts, 43.66.Pn, 43.66 Dc

4.1.2 INTRODUCTION

Dynamic range compression is routinely used in hearing aids to address the limited dynamic range available to hearing-impaired listeners (Moore, 2007). Such compression generally improves audibility and speech intelligibility (Moore, 1996; Jenstad et al., 1999). However, when applied independently to both ears, dynamic range compression can alter interaural level differences (ILDs), which provide important information about acoustic source location (Byrne and Noble, 1998). It is not clear, however, how such alterations in ILDs influence spatial auditory perception. Compression has little effect on the ability of either normal-hearing or hearing-impaired listeners to accurately localize sounds presented in isolation (Keidser et al., 2006; Musa-Shufani et al., 2006). Yet compression can degrade the ability to discriminate small differences in ILD (Musa-Shufani et al., 2006), and can affect normal-hearing listeners' perception of other spatial attributes, such as source diffuseness and perceived movement (Wiggins and Seeber, 2011, 2012). Recent work suggests that independent compression impairs the ability to use spatial cues to selectively attend to a target talker in the presence of other, simultaneous talkers (Kalluri and Edwards, 2007). Motivated by this finding, the current study sets out to further explore whether independent compression interferes with the ability to attend to target speech based on its location in the presence of competing speech, even though it may not adversely affect other aspects of spatial perception. Such a finding can help to resolve the apparent inconsistencies in previous reports of effects, or lack thereof, of compression on different aspects of spatial hearing.

While, intuitively, sound localization is related to the ability to attend to sounds based on their location, accurate localization is neither necessary nor sufficient for predicting the importance of spatial cues in understanding a signal in a mixture of sounds (Noble et al., 1997; Gallun et al., 2008; Schwartz et al., 2012). How dynamic range compression affects spatial selective auditory attention cannot be easily predicted by how it affects localization. Furthermore, many previous studies examining the effect of compression on localization used isolated stimuli. In this case, the stimulus onset is unaffected by compression for some time (depending on the compression time constants), and the "clean" spatial cues at onset may support accurate localization. The compressed ILDs during the ongoing portion of an isolated stimulus also follow a predictable pattern that listeners may be able to learn how to localize (e.g., Bauer, 1966; Shinn-Cunningham et al., 1998). Such factors may help explain why some previous reports find little or no effect of dynamic range compression on localization accuracy for sources in isolation (Keidser et al., 2006; Musa-Shufani et al., 2006).

Spatial acoustic cues such as ILD can be particularly helpful in allowing listeners to attend to and understand a desired sound source when multiple competing sources are present (e.g., Shinn-Cunningham, 2005, 2008; Shinn-Cunningham and Best, 2008). The term "spatial selective auditory attention" refers to cases in which listeners specifically use spatial cues to focus on a desired sound source and mediate competition from distracting sources from other locations (e.g., Ruggles and Shinn-Cunningham, 2011). Spatial separation between competing sound streams also can make it easier to understand a desired sound source by changing the signal-to-noise ratio (SNR) of the signals reaching the ears due to acoustic head shadow (Zurek, 1993); however, such "energetic" factors are often modest compared to effects on spatial selective attention (e.g., Kidd et al., 2005b).

In order to predict how dynamic range compression influences spatial selective auditory attention, it is important to consider the ways in which it alters binaural cues. In general this will depend both on stimulus factors and on compression parameters. A simple dynamic range compression scheme applies linear gain for low-level sounds, and provides progressively less amplification of sounds whose level exceeds a set threshold. Overall, compression that operates independently at the two ears will tend to reduce ILDs, applying a greater gain at the ear receiving the less intense signal. Since many natural stimuli, including speech, fluctuate in level, compression that is applied independently at the two ears will also introduce fluctuations to the ILD of an isolated sound source even if it is stationary in space. The resulting ILD fluctuations are likely to increase source image diffuseness and to cause listeners to perceive moving and split images, even if the source produced constant spatial cues prior to compression (Wiggins and Seeber, 2011, 2012). When multiple sound sources are present in an acoustic scene, the ILD of a target sound source can be altered due to the compressors' response to an unrelated source. In this case, even a target with zero ILD prior to compression can contain fluctuating ILDs; moreover, the ILD fluctuations can occur unpredictably in time and in unpredictable directions. Thus the problems reported by Wiggins and Seeber may be exacerbated in situations where multiple, dynamic stimuli are present, such as speech mixtures.

Dynamic range compression in hearing aids is typically set to operate on timescales anywhere from a few ms to several seconds. If compression is fast (less than 10 ms), then only stimuli that are more or less simultaneous will affect each other's ILDs. Moderate compression speeds on the order of tens of ms could increase interactions between sources because the difference in gain applied to the left and right ears due to a preceding stimulus can affect a target stimulus' ILD, particularly at its onset. If compression is very slow (seconds or more), however, then the response of the compressors will be roughly constant for any given source configuration, resulting in stable spatial cues.

Given these observations, we hypothesized that bilaterally independent dynamic range compression operating on fast-to-moderate timescales would increase the minimum azimuthal separation between competing speech sources required for listeners to effectively attend to a target source. More specifically, when two stimuli are sufficiently close in location, so that ILD fluctuations cause the ILD distributions associated with the two stimuli to overlap, the ability to direct spatial selective auditory attention should suffer and performance should decrease. If, however, the sound sources are sufficiently separated in azimuth such that ILD fluctuations do not cause any such overlap, then compression may not have any effect on spatial selective auditory attention (i.e., the compression-altered ILDs still identify the sources uniquely, with or without these fluctuations). In normal-hearing listeners, the benefit of spatial separation on spatial selective auditory attention increases with azimuthal separation up to about 30°, where it reaches a maximum (Marrone et al., 2008b). Therefore, as the spatial separation of sources increases beyond 30°, we might expect to see decreasing effects of compression.

To test whether compression influences spatial selective auditory attention, we estimated "spatial thresholds," defined as the amount of spatial separation required to achieve a threshold level of performance in a selective attention task, for normal-hearing subjects. Target and maskers were comprised of digits spoken by the same talker. Because of this design, cues such as pitch and semantic context could not be used to focus selective attention, making spatial cues critical for performing the task. We simulated various spatial configurations using head-related transfer functions (HRTFs), preserving realistic spatial cues, including interaural time differences (ITDs), in the resulting stimuli. We compared spatial threshold estimates when amplification was linear (uncompressed), compressive and independent at the two ears, and compressive but linked at the two ears (i.e., left and right ears get an identical gain at each time instant in each frequency channel). The overall stimulus level was set to be similar across conditions. As discussed above, the time constants used by the compressors will influence the compressors' effect on ILDs in the mixture; therefore, Experiment 1 measured spatial thresholds for different combinations of attack and release time constants. Because reverberant energy tends to reduce ILD cues (Shinn-Cunningham et al., 2005), room acoustics are likely to interact with effects of compression on spatial selective auditory attention. Therefore, in Experiment 2, we measured spatial thresholds for different levels of reverberant energy.

4.1.3 METHODS

Stimuli

All source stimuli were digits (0-9) recorded in our laboratory by a male talker in a sound-isolated booth with perforated metal walls, ceiling, and door, and a carpeted floor. The digits were recorded individually, so there was no co-articulation, with a sample rate of 16 kHz. Ten instances of each digit were recorded; when selecting any digit, one of these ten was selected at random.

For each trial, three sequences of four digits were created: one target sequence and two masker sequences. The digits for each sequence were selected at random with the restriction that no digit was presented at the same temporal position in any of the three sequences (for example, if the first digit of the target was 1, then neither of the maskers could also have first digits of 1). We randomly staggered the onsets of each of the digits in the target and two masker sequences. Specifically, for each temporal position (1-4) in the sequences, one of the digits from the three sequences was selected randomly to start first; one of the remaining two digits was randomly selected to start 150 ms after the onset of the first digit, and the final remaining digit started 300 ms after the first digit. As a result of this temporal randomization, the digits within each sequence were not isochronous; the inter-digit interval between two digits in each sequence could be between 300 and 900 ms.

The choice of randomly staggering the onsets of target and maskers was motivated by two ideas. First, synchrony of competing sounds reduces the benefit of spatial separation of the sounds (Best et al., 2011); therefore, staggering the onsets was likely to maximize effects in cases where spatial cues were useful. Second, in natural settings, listeners typically hear a target amidst maskers whose amplitude fluctuations are independent of those in the target, and this independence affects the way in which the maskers will alter that target's ILD. As noted below, we tested performance when maskers were placed symmetrically on either side of a central target. If both maskers were played synchronously without staggering their onsets, the overall levels at the ears would be unnaturally similar at all moments. This in turn could lead us to underestimate any effects of independent compression on the influence of spatial cues on performance.

Spatialization

Because the target and maskers were selected from the same recordings all using the same talker, the target could not be distinguished from the maskers using voice or pitch cues; only spatial cues enabled the listener to select the target stream from the ongoing source mixture. The target sequence was always simulated from straight ahead of the listener, digits from one masker sequence were simulated from locations left of midline, and digits from the other masker sequence were simulated from locations right of midline. To help listeners focus attention on the correct sequence, one second before the start of the sequences listeners were cued by a presentation of the target sequence's first digit played in isolation, simulated from straight ahead. In other words, each trial consisted of a cue digit presented from straight ahead, followed by a 1 s pause, followed by a presentation of three nearly simultaneous streams (left masker, center target, and right masker). Because the first target digit (after the 1 s gap) was the same as the cue digit, responses to the first digit were not included in analyses of performance (post-hoc analysis confirmed that subjects were at ceiling in reporting the first, cued target digit).

To simulate realistic spatial positions with normally occurring spatial cues, sources were convolved with spatial filters. Experiment 1 used head-related impulse responses (HRIRs) that were recorded in an anechoic chamber, sampling every 5° on the azimuthal plane (Gardner and Martin, 1994). Experiment 2 used binaural room-impulse responses (BRIRs) recorded in our laboratory from the ears of the first author, following the procedures described in Shinn-Cunningham et al. (2005). These recordings were made in a 9×5×3.5 m room with a carpeted floor and acoustic tile ceiling; the room had a reverberation time of T_{60} = 650 ms. We recorded BRIRs every 5° from -30° to +30°, as well as at ±40°, 50°, 60°, 70° and

90°. Three different levels of reverberant energy were tested in Experiment 2. The "Reverberant" condition used the measured BRIRs. These BRIRs were modified to create "Intermediate" and "Anechoic" conditions by locating the first reflection of each recorded impulse response manually (separately for left and right recordings). The BRIR was then temporally windowed to isolate the direct sound impulse response (up to the onset of the first reflection) and reverberant energy (from the first reflection onwards). For the Intermediate condition, the reverberant component of each BRIR was attenuated by 6 dB, and then added back to the direct sound. For the Anechoic condition, only the direct sound portion of the BRIR was used. Subjective listening confirmed that these intermediate and pseudo-anechoic BRIRs produced spatialized sources without noticeable artifacts.

As noted above, the target sequence was always simulated from 0° azimuth. The masker positions varied as described in the adaptive procedure, below; however, the two masker digits presented at a given time were symmetrically positioned around midline (e.g., if one masker was at +30°, the other masker was at -30°). This symmetric masking setup was adopted to limit the influence of differences in the long-term SNR between the mixtures reaching the two ears due to head shadow (Marrone et al., 2008b). Nonetheless, short-term SNR will still fluctuate at the two ears even when maskers are symmetrically placed (especially due to the onset staggering we employed); this in turn can affect the ability to hear the target. Moreover, as discussed above, these fluctuations likely interact with the dynamic amplification used in a given condition and thereby may affect how spatial separation influences performance.

Compression

After the binaural speech mixtures were generated, they were put through a simulation of hearing-aid compression. Compression operated independently on 16 frequency bands, equally spaced on the ERB scale from 100 Hz to 8 kHz (Glasberg and Moore, 1990). The compression threshold within each band was set to the level received in that band from speech that was presented at an overall level of 50 dB SPL (estimated used a 5-s token of speech-shaped noise). Linear gain was applied below this threshold. For all conditions employing non-linear compression, a compression ratio of 3:1 was always used above this threshold. This compression ratio is a relatively extreme value compared to typical hearing-aid fittings, and represents a "worst-case" scenario in order to reveal any potential effects of compression. The linear gain below the compression threshold was set so that the 0 dB point (i.e., the level within a band at which 0 gain was applied) was the in-band level of 70 dB SPL speech. The level of each individual digit was set to 70 dB SPL before filtering and compression; therefore, the overall stimulus level was roughly equivalent regardless of the compression condition.

In both Experiments 1 and 2, we used three compression conditions: linear processing, independent compression, and linked compression. For linear processing, the same multiband compressor algorithm was used as in the other conditions, but with the compression ratio set to 1:1 (ensuring that any effects we observed were not due to the multiband analysis / resynthesis, but due to the compression). For linked compression, in any given 8-ms time window, the gain applied to both the left and right ear signal in a particular frequency band equaled the minimum of the gains that would have been applied to the two signals when the compression was independent. The linked compression condition therefore had a slightly lower binaural level than the independent compression condition; whenever a non-zero ILD was present in a compression band, the ear with the less intense signal received less gain in the linked compression condition compared to in the independent compression condition. In Experiment 1, the "Fast" condition used attack and release times of 11 and 82 ms, respectively; (ANSI, 2003), and the "Slow" condition used attack and release times of 48 and 730 ms, respectively. Experiment 2 used only the fast attack and release times. In all cases, the compression scheme estimated power within each band using 8-ms time windows, then smoothed this power estimate with the appropriate attack and release time constants before determining the amount of gain to apply.

Adaptive procedure

We designed an adaptive procedure to estimate subjects' spatial threshold, defined as the separation needed between target and maskers to obtain threshold-level performance. Specifically, the lateral position of the symmetrically placed maskers was adaptively varied until the percentage of target digits correctly reported reached threshold. In Experiment 1, the adaptive procedure tracked 67% correct using a weighted up-down procedure (Kaernbach, 1991): the masker position was decreased by 5° after each correct response and increased by 10° after each incorrect response. In Experiment 2, the 50%-correct threshold was found using a 1-up 1-down procedure (Levitt, 1971). Note that in Experiment 2, the BRIRs were not spaced evenly throughout the azimuthal plane. Therefore, the adaptive track increased or decreased the lateral positions of the maskers by one azimuthal sample (5° for sources near midline; 10° for more lateral locations). In both experiments, an adaptive run continued until 12 reversals were recorded; spatial threshold for that run was then estimated as the median of the last 8 reversals.

One concern with this adaptive procedure is that the distribution of source positions being heard can alter spatial sensitivity over relatively short time periods of time, essentially optimizing spatial resolution over the expected stimulus range (Shinn-Cunningham et al., 1998; Dahmen et al., 2010). As an adaptive run converges on some nominal value of spatial threshold, the range of presented spatial locations narrows to a limited range around the adaptive threshold. This may allow the listener to achieve better spatial resolution than what we might expect in a typical crowded environment, where sounds come from unpredictable directions (precluding the kind of short-term adaptation that would yield better resolution). We therefore interspersed "probe" presentations (those in which the locations of the maskers were varied to adapt to threshold, as described above) with "fixed" presentations within each trial. Specifically, as described above, each trial consisted of a sequence of four target digits presented with symmetric maskers. For three of these four digits, the symmetric masker digits came from one of three fixed spatial separations: ±15°, ±30°, and ±90°. The remaining digit was the probe digit, whose masker spatial angle was determined by the adaptive procedure. Only the response to the probe digit determined whether to increment or decrement the symmetric masker azimuths of the probe digit in the next trial. The order of the fixed and probe digits were randomized on each trial as follows: (1) one of the two middle digits was randomly designated as the probe digit (reducing the effect of recency and primacy on the subject's probability of obtaining a correct response to the probe digit; e.g., Jahnke, 1965; Ruggles and Shinn-Cunningham, 2011); (2) the first pair of masker digits were always given azimuths of ±30°; and 3) the remaining two pairs of masker digits were randomly assigned fixed azimuths so that one pair was at ±15° and the other at ±90°. Recall that the first target digit is used to cue the subject to the target's location; therefore, listeners are expected to be near perfect at identifying the first target digit. This procedure produced useful observations for maskers at the fixed azimuths of ±15° and ±90° at the cost of having no useful observations at the ±30° separation used for the first pair of masker digits. This procedure therefore not only estimates spatial threshold from the probe digits, but also yields fixed-increment estimates of performance for spatial separations of 15° and 90°.

Subjects first ran at least four adaptive runs as training. This training was conducted in two experimental blocks, each comprised of one adaptive track using linear processing and one adaptive track using independent compression, ordered randomly. Subjects were allowed to perform additional training blocks until they were comfortable with the task. Subjects then ran a block of 12 adaptive runs (four for each compression condition). To reduce any effects of learning over the course of the experiment as a confound in our analysis, compression conditions (independent, linear, and linked) in these 12 runs were randomly ordered subject to the following conditions: (1) runs were always paired so that subjects completed two runs of the same kind in a row, and (2) subjects completed one pair of each of the three compression conditions before encountering a second pair of any condition. Spatial threshold was averaged over the four runs of any given compression condition.

Performance for 15° and 90° maskers was computed from the percent-correct responses to fixed-azimuth digits that were in the middle of the sequence and transformed into rationalized arcsine units (RAU; Sherbecoe and Studebaker, 2004). RAU scores are similar to percent correct scores for values in the range 20%-80%. For more extreme values, RAU values have a greater range than the corresponding percent-correct scores, but have more uniform variance than percent-correct scores.

Task

Subjects were instructed to type in the four digits spoken by the target talker coming from the center, using the midline cue digit (which was identical to the first target digit) to help them focus attention on the target. Subjects were told to guess when they were unsure of any given target digit in a 4-digit sequence so that each other response digit was in the correct position in the sequence. Feedback was provided for every trial; once a listener pressed enter, the correct digits turned green and the incorrect digits were replaced by a display of the correct digits in red. This display would hold for 1.5 s, then the digits would clear and the next trial would start. Subjects were not informed that any aspect of the stimulus was being controlled adaptively, based on performance; similarly, there was no indication of which digits were from the fixed angular separations and which were part of the embedded adaptive run.

Subjects

In total, 39 subjects participated in the experiments described. All subjects, ranging in age from 18-22, had clinically normal hearing (15 dB HL or better) as verified by pure-tone audiometry for frequencies between 250 Hz and 8 kHz. Subjects gave written consent (overseen by the Boston University Charles River Campus IRB), and were paid an hourly wage in compensation for their efforts.

Subjects were permitted to participate in more than one group, but many subjects were not able to commit to more than a few experiment sessions. We therefore analyze each group separately, making only limited inferences across groups. Subjects were excluded from a particular group analysis if they failed to achieve sufficiently high performance when target and maskers were separated by 90°. Specifically, we computed the percentage of correctly identified target digits when the maskers were fixed at ±90°, averaged across all compression conditions. In most conditions, only those subjects who achieved above 70% correct on these ±90° trials were included; the only exception was the Reverberant group in Experiment 2 (as described below). In all groups except the Reverberant group, recruitment continued until at least 10 subjects met the inclusion criterion. In Experiment 1, one subject was excluded from each of the Fast and Slow subject groups. In Experiment 2, four subjects were excluded from the Anechoic group, and no subjects were excluded from the Intermediate group. Seventeen subjects performed the task in the Reverberant group; of these 17, only 7 met the 70% criterion. Therefore, for this group we relaxed our criterion to include subjects who scored at least 60% correct on the average of the fixed ±90° trials (13 subjects). Table 4.1 summarizes how many subjects performed each condition.

	Spatial filters	**Condition**	**Attack / release time constant**	**Subjects tested**	**Subjects included**
Exp. 1	Gardener and Martin, 1994	Fast	11 / 82 ms	11	10
		Slow	48 / 730 ms	11	10
Exp. 2	Recorded in classroom	Reverberant	11 / 82 ms	17	13
		Intermediate (reverb -6 dB)	11 / 82 ms	10	10
		Anechoic (direct sound)	11 / 82 ms	14	10

Table 4.1 – Summary of the groups tested

4.1.4 RESULTS

Spatial thresholds

In all groups, spatial thresholds were typically below 30°, consistent with previous reports using a similar symmetrical masking paradigm (Marrone et al., 2008b). In Experiment 1 (Figure 4.1, top), in which the threshold for 67% correct performance was estimated, thresholds were typically between 10 and 30°. Note that the linear processing condition was identical for both Fast and Slow groups, as no dynamic range compression was applied. Consistent with this, these groups showed similar spatial thresholds for this condition, suggesting that this metric was similar across the two different subject groups tested. In Experiment 2 (Figure 4.1, bottom), the 50% correct performance level was estimated. In the Reverberant group, the 50% thresholds were largest (around 20°), consistent with the idea that utility of spatial cues is reduced in reverberant environments (Nabelek and Pickett, 1974; Marrone et al., 2008c; Ruggles and Shinn-Cunningham, 2011), although we also had to relax our inclusion criterion to 60% correct for this group, which makes it especially difficult to directly compare groups. Spatial thresholds for the Intermediate and Anechoic groups tended to fall between 10 and 20°, with no subjects having spatial thresholds greater than 25°. These values suggest that ceiling effects may limit the observable differences in spatial thresholds across compression conditions in Experiment 2.

Figure 4.2 plots within-subject differences in spatial thresholds relative to the spatial threshold in the independent condition (which we hypothesized would be largest, due to ILD fluctuations and image diffuseness). For all groups using fast compression (Fast condition in Experiment 1 and all three conditions in Experiment 2), subjects tended to perform worse with independent compression than for either linear processing or linked compression. This produced negative spatial threshold differences in Figure 4.2, consistent with our hypothesis. These differences were small, generally under 10°. Because little is known about the distribution of spatial thresholds across subjects, we used a directional

Wilcoxon sign-rank test, a non-parametric test of significance, for these within-subject differences. In the Fast group (Experiment 1, top panel of Figure 4.2) and the Reverberant group (Experiment 2, bottom panel of Figure 4.2), the differences in spatial threshold relative to independent compression were significant for both linked compression ($p<0.005$ for both groups) and linear compression ($p<0.01$ for Fast, $p<0.05$ for Reverberant). The Intermediate and Anechoic groups tested in Experiment 2 showed the same trends, but the differences in spatial threshold for linear processing and linked compression relative to independent compression were not statistically significant ($p>0.05$ for all comparisons). As mentioned above, smaller differences in these two groups may have resulted from ceiling effects, making it difficult to detect a significant effect of compression. We therefore performed a post-hoc analysis, pooling together the Intermediate and Anechoic groups to increase sample size. This post-hoc test supports the notion that independent compression caused a real, albeit small, decrease in performance ($p<0.05$ for both linear processing and linked compression). For the Slow group of Experiment 1, there was not even a trend for the spatial threshold to be larger in the independent compression condition compared to the linear or linked conditions, suggesting that slower compression speeds may alleviate the detrimental effect of fast compression on selective attention performance.

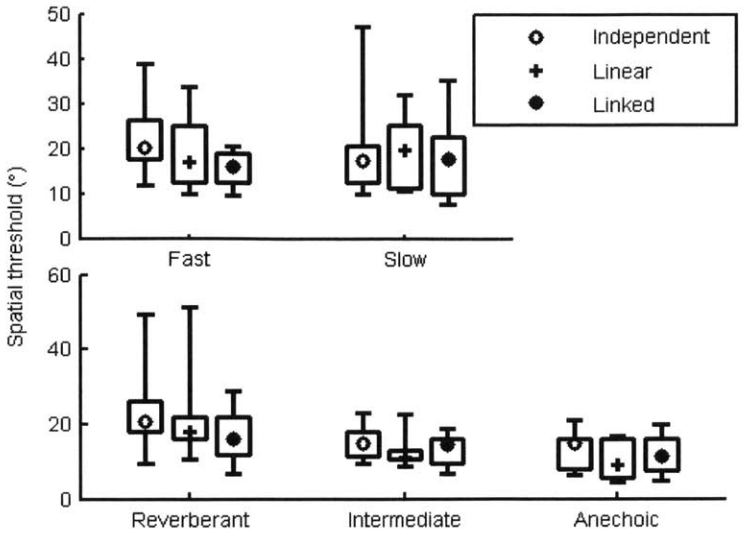

Figure 4.1 Quartiles of subjects' spatial thresholds in Experiment 1 (top) and Experiment 2 (bottom), for independent (open circles), linear (plusses), and linked (filled circles) compression. Symbols mark the medians of each data set, boxes indicate the inter-quartile range, and whiskers surround the full range of results. N=10 for each condition except Reverberant, in which N=13.

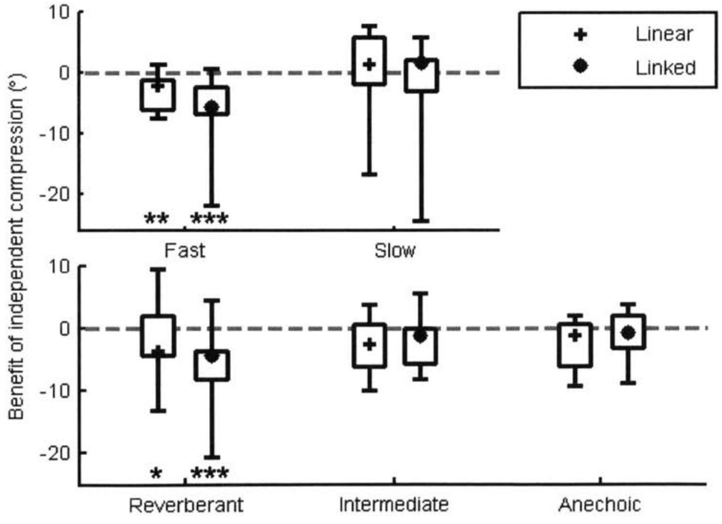

Figure 4.2 Quartiles of within-subject differences of spatial thresholds for linear (plusses) and linked compression (filled circles), both relative to independent compression. Negative values indicate smaller thresholds (better performance) compared to independent compression. Statistical significance is assessed by a directional Wilcoxon signed-rank test, and indicated by a * for $p<0.05$, ** for $p<0.01$, and *** for $p<0.005$.

We also analyzed whether spatial threshold changed over the time course of the experiment. For each condition, we computed the difference in spatial threshold averaged over the first two runs and over the final two runs. We found no differences between results in the first and final pairs of runs for any individual group or compression condition; we also found no differences when pooling across groups, compression conditions, or both groups and compression conditions (Wilcoxon sign-rank test, $p>0.05$ for all comparisons). These results suggest that performance was stable over the course of our experiment. We therefore combined all four runs of any given condition to compute mean spatial threshold.

Fixed-azimuth performance

While the focus of our data collection was to estimate spatial thresholds, our methods also allowed us to perform post-hoc analysis on subjects' performance for digits with masker azimuths fixed at 15° and 90°. Performance for fixed digits was computed (in RAU) separately for digits that occurred in the middle of the four-digit target sequence and digits that occurred at the end of the sequence. No consistent within-subject differences were found in performance between middle and end digits (t test; $p>0.05$). All remaining analyses were therefore performed by combining results for the fixed-azimuth digits regardless of where in the sequence they occurred. The exact number of trials included in this computation for a given condition varied due to the adaptive procedure employed; across all conditions and groups, individual subjects completed between 78 and 130 trials for any given compression condition, with a majority of subjects performing between 90 and 110 trials per condition.

Figure 4.3 plots the RAU scores for fixed-azimuth maskers. The dashed line shows the threshold performance level estimated by the adaptive procedure (corresponding to 67% correct in Experiment 1 and 50% correct in Experiment 2). Averaged across groups, performance was 17.2 ± 4.2 RAU better for 90° maskers than for 15° maskers (mean ± standard deviation; however, in some groups a few subjects were excluded from all analysis based on poor performance for 90° maskers). Note that for all subject groups in both experiments, except the Reverberant group, RAU scores tended to be above this dashed line, even for 15° maskers; therefore, spatial threshold estimates made by fitting these data would tend to produce values less than 15°. Spatial thresholds estimated by the adaptive procedure, on the other hand, often fell near or above 15°, indicating that these two methods produce slightly different estimates of performance. This difference is not surprising with near-ceiling performance, as the adaptively varied azimuth was bounded by ceiling (0°), producing a biased estimate of spatial thresholds. Yet, even despite this bias, which should limit observable differences in the spatial thresholds across conditions, both spatial thresholds and fixed-azimuth performance measures produce similar effects.

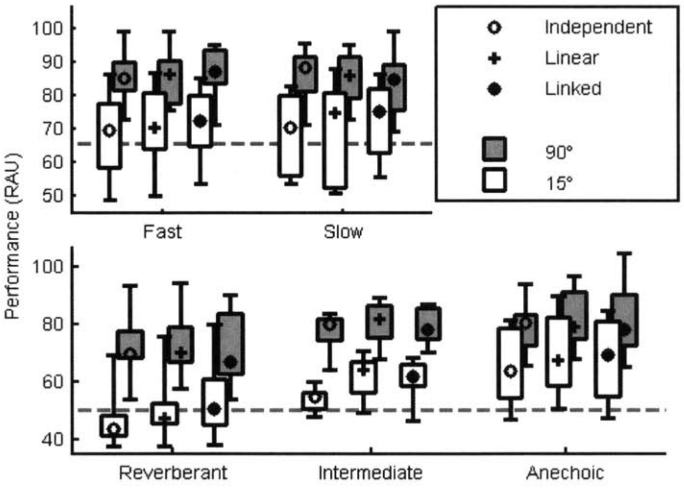

Figure 4.3 – Quartiles of RAU scores for fixed-azimuth digits with maskers at 15° (white boxes) and 90° (grey boxes) for linear, independent compression, and linked compression (plusses, open circles, and filled circles, respectively). Grey-dashed lines represent the spatial threshold performance point estimated by the adaptive procedure for the corresponding groups (~65.5 RAU and 50 RAU for the TIME groups and the REVERB groups, respectively).

Figure 4.4 plots within-subject differences in RAU scores for independent compression relative to linear and linked compression for the 15° separation (90° results are not plotted for visual clarity since none of these differences were significant). The effect of compression was assessed separately for 15° and 90°

maskers using one-sided t-tests on the differences between linear and linked compression relative to independent compression, all in a within-subject design. These differences revealed a distinctive pattern. There were no significant differences between compression conditions for 90° maskers for any group in either Experiment 1 or Experiment 2. However, for 15° maskers, all groups in Experiment 2 had better performance for linear processing than for independent compression (directional t-test: $p<0.05$ for Reverberant, $p<0.01$ for Intermediate and Anechoic). Moreover, all groups in both Experiment 1 and Experiment 2 had better performance for linked compression than for independent compression (directional t-test: $p<0.05$ for Fast group in Experiment 1, $p<0.01$ for all other groups). The size of this effect varied across subjects, with a mean effect size across all groups of 4.0 and 4.7 RAU for linear processing and linked compression, respectively, vs. independent compression. In Experiment 2, the mean difference was larger for the Intermediate condition than for the other two conditions. While our experiment design does not support direct across-group comparisons, the tendency for the Intermediate condition to reveal the largest effects of independent compression may reflect the fact that subjects' spatial thresholds were more consistently close to 15° in this condition (see Figure 4.1), resulting in relatively more sensitive measures of performance compared to conditions where subjects had thresholds greater than 15° (e.g., Reverberant) or smaller than 15° (e.g., Anechoic).

The timing of target and masker digits within each of the four temporal positions of the digit sequences was randomly staggered by 0, 150, or 300 ms. While this was not the focus of our study, post-hoc analysis revealed an effect of this temporal staggering on performance, with target identification tending to be better when the target was given a 150 or 300 ms delay compared to when it was given no delay. However, this effect did not interact with any of the effects of compression, which was the focus of our study. We are currently investigating this effect further.

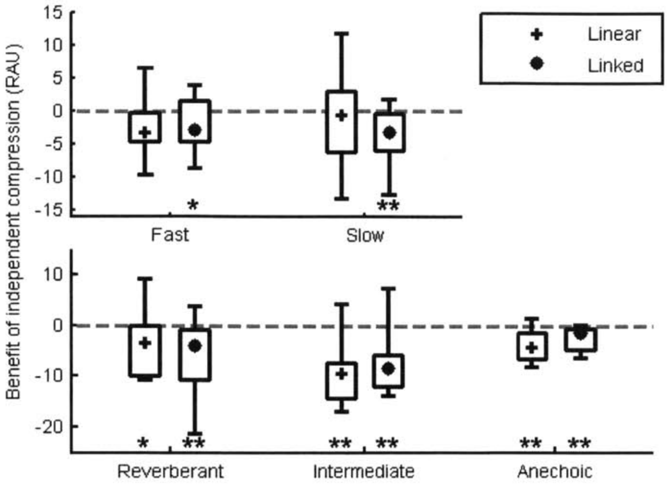

Figure 4.4 – Quartiles of within-subject differences of RAU scores for linear (plusses) and linked compression (filled circles) relative to independent compression for 15° maskers. Positive values indicate better performance compared to independent compression. Statistical significance was assessed by pairwise t-tests, and is marked with a * for p<0.05 and ** for p<0.01. Differences for 90° maskers were not significant for any group and are omitted for visual clarity.

Error type analysis

We analyzed the types of errors made by subjects on each individual digit when maskers were fixed at 15°. Specifically, we categorized errors into "switch" errors, in which the reported digit was present in the reported temporal position, but came from one of the two masker sequences, and "drop" errors, in which the reported digit was not present in the reported temporal position in any of the three sequences. For this analysis, we included all tested subjects regardless of their average performance for 90° maskers. Due to the random temporal staggering we imposed on the digits, it is possible that subjects may have mistaken masker digits for target digits not in exactly the same temporal position (for example, the third target digit might be separated by ±300 ms both from the third digit in a masker sequence and also the second digit in a masker sequence). Therefore, some small percentage of errors that truly result from improper selection across temporal positions within a digit sequence may be incorrectly counted as "drop" errors in this analysis. Nevertheless, if errors resulted entirely from random guessing, then switch errors should make up roughly 22% of all errors (2 masker digits out of 9 possible non-target digits). However, even with the potential for undercounting switch errors, they made up 79% ± 11% of all errors (mean ± standard deviation across all subjects). Moreover, each individual subject made significantly more switch errors than would be expected by chance (binomial test, $p \ll 0.0001$). This result indicates that most errors were made not due to lack of intelligibility of the digits, but instead due to a failure to select the correct digit among the three digit sequences (see also: Kidd et al., 2005; Ihlefeld and Shinn-Cunningham, 2008; Ruggles and Shinn-Cunningham, 2011).

In all groups, the overall error rate for 15° maskers with independent compression was greater than for either linear processing or linked compression by an average of 3.2 and 4.7% of trials, respectively (mean error rates were 42.1, 38.9, and 37.3% of trials for the three conditions, respectively). Even though most errors were switch errors, these differences between conditions might depend on the pattern of drop errors, rather than switch errors. To explore this, we compared the percentage of switch and drop errors for linear and linked compression to the percentage for independent compression on an individual subject basis. Figure 4.5 plots both the overall error rates for switch and drop errors (left panel), and within-subject differences in the switch and drop error rates for linear processing or linked compression compared to independent compression (right panel). Drop errors constituted less than 10% of the responses in all conditions, while switch errors occurred on almost 1/3 of all responses (see left panel of Figure 4.5). On an individual subject basis, drop errors rates were statistically the same when comparing independent compression to linear and to linked compression (average differences of 0.6% and -0.2%, respectively, $p>0.05$ for both using a one-tailed t-test on RAU-transformed values; shaded boxes in the right panel of Figure 4.5). Switch error rates, however, were significantly higher for independent compression than for either linear processing or linked compression (an average increase of 2.5 and 5.0% of trials; one-tailed t-test on RAU transformed values, $p<0.001$ and $p \ll 0.0001$, respectively; white boxes in the right panel of Figure 4.5). These results suggest that independent compression in our experiment impaired performance by interfering with selection of the proper source, not by degrading overall signal intelligibility.

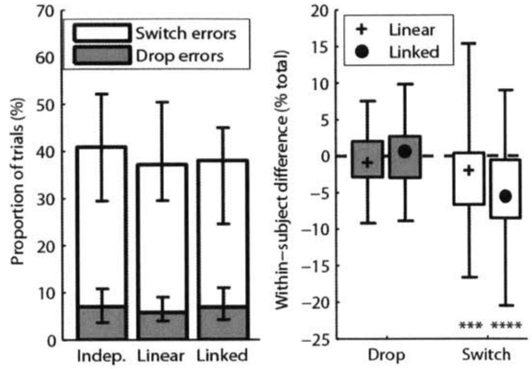

Figure 4.5 – Left: Error rates for switch errors, in which subjects reported a masker digit, and drop errors, in which subjects reported a digit not in either target or masker sequences. Bars indicate median error rate across all subject in both experiments (N=63); error bars represent inter-quartile range. Right: Within-subject differences in error rates for switch and drop errors for linear and linked compression relative to independent compression. Negative values indicate lower drop or switch errors relative to independent compression. Symbols indicate the median across all subjects, boxes show the interquartile range, and whiskers show the full range. Statistical significance is assessed by Wilcoxon sign-rank tests, and is marked by *** for p<0.001 and **** for p<0.0001.

4.1.5 DISCUSSION

Fast, independent compression elevates spatial thresholds in normal-hearing listeners

In all groups using fast compression, average spatial thresholds were larger (a greater spatial separation was needed for listeners to perform at threshold) for independent compression compared to linear or linked compression (Figure 4.2). Differences were modest on average (around 5°), with large inter-subject differences. It is yet unclear what practical effect on communication and listening effort such differences may imply for real-world settings, particularly using more realistic and complex hearing aid compression schemes; nevertheless, these results support the idea that independent compression interferes with spatial selective auditory attention and increases the difficulty of attending to a target sound source amongst competing sources. Such listening situations are considerably more effortful for hearing-impaired than for normal-hearing listeners (Gatehouse and Noble, 2004; Edwards, 2007). If effects like those demonstrated here are also seen in hearing-impaired listeners (see Sec. IV C), they could have an important impact on the ability of hearing-aid users to communicate in everyday settings.

Effects of slow compression in our data are less clear than those for fast compression. Some previous studies support the idea that slow compression has a smaller effect on spatial perception than faster compression. For instance, ILD sensitivity in quiet is more adversely affected by compression using faster time constants compared to slower time constants (Musa-Shufani et al., 2006). Slow compression also yields better performance than fast compression in hearing-impaired individuals asked to report target sentences presented with spatially separated speech maskers (Moore et al., 2010). Slower compression will generally result in smaller ILD fluctuations, as the gain changes in either ear are relatively less affected by instantaneous fluctuations in signal power and relatively more by longer-term average power at the two ears. Slower fluctuations in ILD can also be more easily tracked by the binaural system compared to fast fluctuations, which increase image diffuseness (Grantham and Wightman, 1978; Culling and Colburn, 2000). Slow compression may also provide longer "clean" ILDs at stimulus onsets that dominate spatial perception (Freyman et al., 1997; Stecker and Hafter, 2002); these onsets can are also be further enhanced by dynamic range compression (Verschuure et al., 1996), which may increase onset dominance.

Spatial and non-spatial factors can contribute to effects

In addition to the elevation of spatial thresholds, our results support the idea that spatial attention plays a role in our paradigm: we found that performance was significantly worse with fast, independent compression than with linear processing or linked compression only when target and maskers were separated by a small angular separation (15°). Independent compression did not affect performance for large azimuthal separations (90°). We suggest that increased image width and diffuseness caused by independent compression (Wiggins and Seeber, 2012) only affect spatial selective attention if competing sources are sufficiently close to each other in azimuth such that these effects cause confusion about whether a particular sound is from the target or a masker. A more thorough analysis of the acoustic effects of compression on the spatial cues available to normal-hearing and hearing-impaired listeners can lend further insight into this idea, and is one focus of our future research.

Differences across the compression conditions were driven by “switch” errors, in which subjects selected one of the masker digits, further supporting the idea that compression interfered primarily with source selection rather than with speech intelligibility. Similar results can occur even in diotic mixtures (increased "reversals"; Stone et al., 2009), indicating that overall cognitive load, and not necessarily spatial factors, may also contribute to our results. In considering this possibility, it is important to note that the previous study that found increased reversals for diotic mixtures used a cognitively demanding task in which subjects divided attention between two simultaneous streams (see also: Gallun et al., 2007; Best et al., 2010) and responded only after performing an unrelated, visual distractor task (further increasing cognitive load). In addition, in that study, listeners could report the content of the two streams in any order for them to be counted as correct. It is plausible, then, that the reversals reported in this earlier diotic study might not reflect a failure of incorrect selection, but instead may represent a memory failure in which, for example, listeners could recall the words spoken but were unable to accurately bind the identified speech tokens to the correct talker identity (see also: Treisman and Gelade, 1980; Woods et al., 1998). In contrast, our results show a clearer failure of selection using a task that requires subjects to attend to only a single source (comprised of only four digits, which should not stress working memory) and without any secondary task. As correct selection could only be done in our task using spatial cues, and as switch errors were reduced when compressors were linked across the two ears, we argue that the main effect of independent compression in the current study came about from failures of spatial selection, rather than due to non-spatial affects.

Nevertheless, non-spatial factors, including reduced spectral and temporal modulation and across-signal modulation correlation (Stone and Moore, 2003, 2007; Stone et al., 2009), may have also played a role in our results. For example, it may be that only when maskers were located at 15°, but not at 90°, was our task sufficiently challenging to show effects on performance due to fast, independent compression. In addition, we linked the left and right compressors to preserve ILDs; however, improved performance in this condition relative to independent compression may also be due to a lower "effective compression ratio" (Stone and Moore, 1992) rather than due to preservation of ILDs. Future work could be done to clarify this difference by using a higher target compression ratio in the linked compression condition so that the effective compression ratios are equal.

Compression may affect ILD utility differently in hearing-impaired listeners

Our results demonstrate that fast independent compression with a high compression ratio can elevate spatial thresholds in normal-hearing listeners. However, the same manipulations may influence hearing-impaired listeners differently. The acoustic features that allow normal-hearing listeners to segregate the target from the maskers and selectively listen to the target may not be fully available to hearing-impaired listeners (e.g., see discussion in Shinn-Cunningham and Best, 2008). Hearing-impaired listeners may have reduced sensitivity to binaural cues or use different listening strategies, so that the practical consequences of compression on the influence of spatial perception are limited. For example, if binaural processing is compromised, further corruption of ILD by compression may have no noticeable effect on selective auditory attention.

Another issue is that healthy ears naturally compress acoustic inputs through the nonlinear amplification of the cochlea. In contrast, hearing impairment is often accompanied by a loss of compressive cochlear amplification (Moore, 2007). Hearing-aid compression may approximately restore the kind of compressive amplification that occurs naturally in a healthy ear. Given this, the reduction of ILDs caused by compression may in fact restore "normal" neural representations of ILD in many individuals with hearing loss, leading us to overestimate the effects of compression by testing normal-hearing listeners. However, even if the restoration of normal ILDs improved sound localization accuracy, it would not necessarily improve the ability to use spatial cues to direct attention (Noble et al., 1997; Gallun et al., 2008; Schwartz et al., 2012). We argue that in order to correctly select a target talker using spatial cues (or any other cues), listeners only need to be able to distinguish the target from the maskers using one or more perceptual features, such as location. If hearing-aid compression reduces ILDs, even if it approximately restores average neural ILD representations, it still reduces the difference between target and masker ILDs. This is likely to make it more difficult to use ILD to distinguish the target from the maskers in an acoustic mixture.

Additionally, ILD fluctuations, especially those caused to an attended source by other, unattended sources, may also affect spatial selective auditory attention in hearing-impaired listeners, even if the compression restores neural ILDs to something more like "normal." Previous data suggest that ILD fluctuations due to the dynamic nature of compression, more than the overall ILD reduction, are primarily responsible for perceptually relevant effects in normal-hearing listeners (Wiggins and Seeber, 2011, 2012). It is reasonable to suspect, then, that the dynamics of hearing aid compression may have deleterious effects on the ability of hearing aid wearers to use spatial cues to attend to a desired sound source. Further research with hearing-impaired listeners, using a more representative range of compression settings, can help clarify the practical consequences of the effects being explored here.

Symmetric spatial configuration may have reduced observed differences

In our experiment, we chose to place maskers symmetrically about midline. In retrospect, this choice may have led us to underestimate the possible size of effects. We found that performance was impaired when the maskers were close (15°) to the target, but not when they were far (90°). For close maskers, the magnitude of ILDs in the acoustic mixture is small relative to those present when maskers are far; consequently, the effect of the maskers on the target ILD is relatively small. If performance is impaired due to ILD fluctuations imposed on the target by distracting sources, then we would expect to see a relatively larger effect by, for example, placing one masker close to the target and the other masker at a more lateral location. Such an experiment would also help clarify the contributions of spatial and non-spatial effects of compression.

4.1.6 CONCLUSIONS

Fast, independent binaural compression impairs the ability of normal-hearing listeners to select a desired target from a mixture containing spatially separated maskers. Linking left- and right-ear compression so that the gain applied to the two ears is the same at each time instant preserves normally occurring spatial cues, and restores the ability of normal-hearing listeners to successfully hear out a target stream based on its location. For large spatial separations, performance is relatively good, and not affected by any of the compression schemes tested, consistent with the idea that even when spatial images are made more diffuse, sufficient spatial separation allows for the successful selection of the target. These results highlight the importance of considering a variety of spatial configurations when assessing binaural listening performance rather than using only a single, relatively large separation. Effects of slower compression on performance are less clear. Further investigation should be conducted to reveal if similar detrimental effects on spatial selective attention occur in hearing-impaired listeners, and, if so, whether such effects may exacerbate the problem of increased listening effort in noisy situations experienced by many such listeners.

4.1.7 ACKNOWLEDGEMENTS

Many thanks to the researchers at Starkey Hearing Research, Berkeley, CA, including Sridhar Kalluri, Olaf Strelcyk, Jing Xia, Brent Edwards, Nazanin Nooraei, and Joyce Rodriguez for helping to create the original hypotheses that drove this work and for creating a better understanding of the practical context and implications of this research. Thanks also to Tim Streeter and Scott Bressler for setting up and running the HRTF measurements, and to Justin Fleming for work contributing to the analysis of temporal offsets on performance. This work was supported by Grant NIDCD ROI DC009477.

4.2 Dynamic range compression effects spatial selection of speech sounds in normal-hearing listeners

Andrew H. Schwartz

Harvard / Massachusetts Institute of Technology, Speech and Hearing Bioscience and Technology Program, Cambridge, Massachusetts 02139

Barbara G. Shinn-Cunningham[3]

Center for Computational Neuroscience and Neural Technology and Biomedical Engineering, Boston University, Boston, Massachusetts 02215

Running title: Spatial attention with bilateral compression

---

[3] Author to whom correspondence should be addressed. Electronic mail: shinn@cns.bu.edu

4.2.1 ABSTRACT

Dynamic range compression operating independently between the two ears impairs the ability of normal-hearing subjects to identify target speech in a mixture [Schwartz and Shinn-Cunningham, J. Acoust. Soc Am. 2013]. This effect may be due to the corruption of interaural level differences (ILDs) or an increase in cognitive effort required to perform the task. Here, normal-hearing listeners were presented with speech mixtures in which only interaural time differences distinguished the target from maskers. Average performance was consistent with previous results. However, there was no effect of compression, suggesting that compressed ILDs were responsible for impaired performance in the previous study.

PACS numbers: 43.66.Ts, 43.66.Pn, 43.66 Dc

4.2.2 INTRODUCTION

Hearing aids often employ dynamic range compression to address the limited dynamic range available to hearing-impaired listeners (Moore, 2007). This compression is often applied independently at each ear. Unfortunately, such compression can alter interaural level differences (ILDs), which are a valuable spatial cue allowing listeners to direct attention to a target speaker in noisy acoustic mixture. Results from recent studies over the last decade have come to different conclusions about the effects of compression on spatial hearing (c.f. Keidser et al., 2006; Musa-Shufani and Walger, 2006; Wiggins and Seeber, 2012).

Recently, we showed that fast, independent compression could negatively affect a normal-hearing listener's ability to attend to target speech in a spatially separated mixture (Schwartz and Shinn-Cunningham, 2013). Specifically, we asked subjects to attend to target digits coming from midline in the presence of a competing pair of masker digits coming from azimuths to the left and right. When compression was fast (11 / 82 ms attack / release time constants; ANSI, 2003) and independent at either ear, listeners were more likely to erroneously report one of the masker digits rather than the target digit compared to when linear processing was used. Linking the compressors to provide identical gain at both ears at each time instant (preserving ILDs) restored performance. However, this effect was only present when masking digits were close to the target (15°), not when they were far (90°), showing that once spatial separation was "large enough," the effects of independent compression did not interfere with focusing spatial selective attention. Consistent with this observation, listeners required a greater degree of spatial separation to reach threshold performance when compression was independent.

It is possible that the observed effects were due to decreased precision of spatial images with independent compression (Musa-Shufani et al., 2006; Wiggins and Seeber, 2012), resulting in confusion between sources. An alternative explanation, however, is that compression simply increased the cognitive load required to perform the task regardless of the spatial configuration. For example, by introducing common modulations between the three speech sources, the digits may have become harder to perceptually segregate (e.g., Stone et al., 2009). Under this view, when the masker digits were at ±90°, the task was sufficiently easy such that no effect of this increased cognitive load was observed. However, when the masker digits were at ±15°, the task was sufficiently challenging that the additional cognitive demand impaired subjects' performance in the task.

Here, we present data from a new experiment using the same digit task, but with, masker digits separated only by an interaural time difference (ITD; i.e., the ILD was zero for all target and masker digits). We argue that if the effect of compression on performance was due simply to increased cognitive load, rather than reduced spatial precision, then compression should have degraded performance in the current experiment, just as in the original study. The results of this study therefore shed further light on the nature of the effects reported in our previous study.

4.2.3 METHODS

Apart from the method used to provide spatial separation to the maskers, our methods, briefly summarized here, were identical to those used in Experiment 2 of Schwartz and Shinn-Cunningham (2013).

Stimuli

All source stimuli were digits (0-9) sampled at 16 kHz, recorded in our laboratory by a male talker in a sound-isolated booth with minimal reverberation. The target and two maskers were sequences of four digits, all from the same talker. At any given time, target and masker digits always differed from each other. Each of the four digits in each sequence was given a random temporal delay of 0, 150, or 300 ms.

To help listeners attend to the correct sequence, the first target digit was played in isolation one second prior to the start of the three nearly-simultaneous sequences.

To distinguish the target from masker digits, the masker digits were given an ITD between -600 and 600 μs. Masker digits were always given equal but opposite ITDs so that they symmetrically flanked the target digit, which always had 0 ITD. Without this ITD separation, no features distinguished the target from the masker digits; subjects would, at best, be guessing between the three digit sequences.

Compression

Speech mixtures were put through a simulation of hearing-aid compression using 16 frequency bands that were equally spaced on the ERB scale (100 Hz to 8 kHz; Glasberg and Moore, 1990). The compression threshold within each band was set to the level received in that band from speech that was presented at an overall level of 50 dB SPL. Linear gain was applied below this threshold such that all digits were presented at roughly 70 dB SPL. The compression ratio was 3:1.

We used three conditions in separate blocks: linear processing (compression ratio set to 1:1), independent compression, and linked compression. For linked compression, at any given time, both left and right-ear compressors applied identical gain to each signal; this gain was chosen as the minimum of the gains that would have been applied to the two signals when the compression was independent. Attack and release times were 11 and 82 ms, respectively. The compression scheme estimated power within each band using 8-ms time windows, then smoothed this power estimate with the appropriate attack and release time constants before determining the amount of gain to apply.

Adaptive procedure

Similar to our previous experiments, we used an adaptive procedure that allowed us to converge on an "ITD threshold," defined as the ITD needed between the target and maskers for 50% correct performance, while simultaneously measuring performance with fixed masker ITDs. To do this, we divided the four digit positions within each sequence as follows. The first digit position always had maskers with ±200 μs ITD (recall that this first target digit was played once ahead of the mixture to prime listeners to the target). One of the two middle positions (second or third, chosen randomly) had maskers whose ITD was varied adaptively using a 1-up 1-down rule (Levitt, 1971), using 25 log-spaced ITD values between 2 and 600 μs. The remaining two digit positions had maskers with ITDs of ±100 μs and ±600 μs (chosen randomly without replacement). Each block of trials was run for 12 reversals of the adaptive ITD; the 50% performance threshold was estimated from the median adaptive ITD at the last 8 reversals (see also: Kaernbach, 1991). Subjects first ran four adaptive runs as training, followed by twelve adaptive runs in the main experiment block (four for each compression condition). The runs within the training and main experiment blocks were ordered randomly (see Schwartz and Shinn-Cunningham, 2013 for more details). ITD threshold was calculated as the average over the four main runs of any given compression condition.

Task

Subjects were instructed to type in the four digits spoken by the target talker coming from the center, and to guess whenever they were unsure of a specific digit. Visual feedback was provided for every trial.

Subjects

Twelve subjects participated in these experiments. Subjects ranged in age from 18-21 and had clinically normal hearing (15 dB HL or better) as verified by pure-tone audiometry for frequencies between 250 Hz and 8 kHz. Subjects gave written consent (overseen by the Boston University Charles River Campus IRB), and were paid an hourly wage in compensation for their efforts. One subject was excluded from analyses for failing to achieve 50% correct performance with the maximum target-masker ITD separation (600 μs).

4.2.4 RESULTS

ITD thresholds

ITD thresholds for our subject population, representing the masker ITD resulting in 50% correct performance, are plotted in Figure 4.6A. There is substantial inter-subject variability, with thresholds falling between roughly 100 and 400 μs. For comparison, using an identical task in which the same digit recordings were spatialized using head-related transfer functions (HRTFs), spatial thresholds tended to lie between 10° and 50° except when limited by ceiling effects (Schwartz and Shinn-Cunningham, 2013). Despite this variability across subjects, in our previous study, the differences between fast, independent compression and either linear processing or linked compression, calculated within subject, showed a clear effect of compression: independent compression resulted in larger thresholds (worse performance). By contrast, in the current experiment, even looking within subject, there is no evidence for an effect of compression for either linear processing or linked compression compared to independent compression (see Figure 4.6B; Wilcoxon sign-rank test, $p>0.05$). This negative result suggests that the main effect of compression shown previously is due to changes in ILD; however, the failure to find an effect could also simply reflect the conservative nature of the non-parametric Wilcoxon sign-rank test. A post-hoc, paired-sample t-test also reveal no significant effect ($p>0.05$).

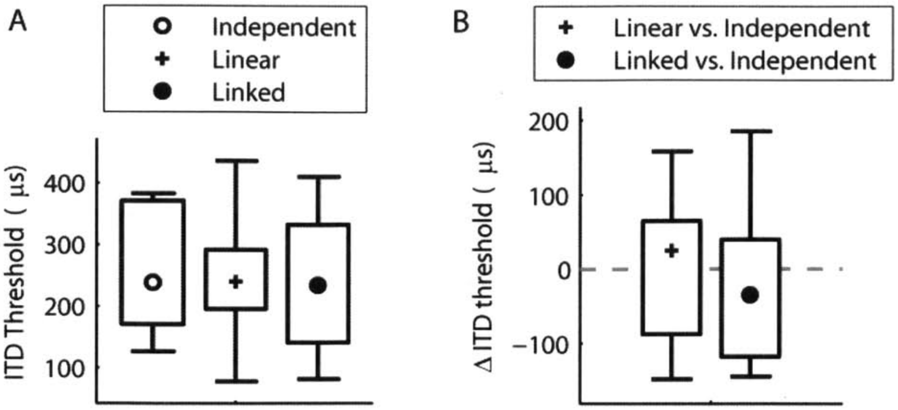

Figure 4.6 – A. 50% performance thresholds, across all subjects. Symbols mark group medians, boxes indicate the interquartile range, and whiskers encompass the full range of results. B. Within-subject change in 50% performance thresholds between either linear processing or linked compression and independent compression. Positive values indicate larger ITD thresholds (worse performance) relative to independent compression; however, no significant were found. Symbols mark group medians, boxes indicate the interquartile range, and whiskers encompass the full range of results.

Fixed-ITD performance

We also looked at overall performance, expressed in RAUs (Sherbecoe and Studebaker, 2004) when maskers were at fixed ITDs. These results are plotted in Figure 4.7A. Performance is better when maskers are given 600 μs ITD (mean 69.8 ± s.d. 9.8 RAU) compared to when maskers are given 100 μs ITD (46.0 ± 9.3 RAU). These scores are overall lower than performance when full-cue HRTFs were used; in that previous study, subjects scored 83.3 ± 9.2 RAU when maskers were located at 90° and 67.9 ± 11.8 RAU when maskers were located at ±15°. Still, for either masker ITD, performance is well above a chance level of 1/3 (directional t-test, $p<<10^{-4}$ for both 100 μs and 600 μs masker ITDs). Note that a chance level of 1/3 is the more conservative value for this test, assuming that subjects were guessing from among the three digit sources rather than guessing digits randomly (which would have led to even lower performance).

In our previous study, when full-cue HRTFs were used to spatialize masker digits, independent compression decreased performance only when maskers were located at ±15°. Overall performance with this spatial separation varied from roughly 40 – 80 RAU, depending on the simulated acoustic environment. Individual differences in performance in the current experiment are shown in Figure 4.7B. Despite covering a similar range of performance values, there is no evidence for any effect of independent compression for either masker ITD value and either linear processing or linked compression compared to independent compression (directional t-test, p>0.05).

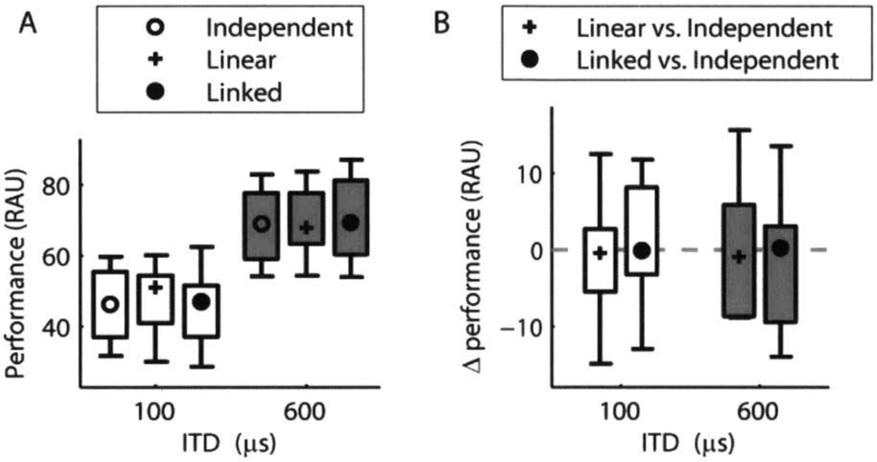

Figure 4.7 – A. Performance, in RAU, for digits when maskers were given ITDs of 100 μs (white boxes) or 600 μs (grey boxes). Symbols mark group medians, boxes indicate the interquartile range, and whiskers encompass the full range of results. B. Within-subject differences in performance between independent compression and linear processing or linked compression. Positive values indicate better performance with independent compression; however, no significant differences were found. Symbols mark group medians, boxes indicate the interquartile range, and whiskers encompass the full range of results.

4.2.5 DISCUSSION & CONCLUSIONS

If independent compression impaired performance in our previous experiments due to non-spatial effects (for instance, increased cognitive load required to segregate the target from the maskers due to common modulations across the signals in the mixture; Stone et al., 2009), we would expect to see a similar effect of compression in the current experiment, where ILDs did not help distinguish between sources. However, we found no evidence for such an effect. Performance was slightly poorer than in the previous experiments using full-cue HRTFs, but still well above chance and within the range of performance values encountered with 15° maskers in the previous study. The effect of compression was only observed in the previous study for maskers located at 15°; therefore, the fact that no effect is observed here argues against the idea that only for maskers located at 15° was the task sufficiently difficult to show an effect of increased cognitive load.

Taken together with the previously reported results, these results offer further support to the idea that fast, independent compression impairs normal-hearing listeners' ability to attend to speech based on its spatial location by corrupting ILDs and degrading the spatial perception of sound.

4.2.6 ACKNOWLEDGEMENTS

This work was supported by Grant NIDCD ROI DC009477. Many thanks go to an anonymous reviewer of our previous manuscript who helped to identify alternative explanations of those data that we address in this manuscript.

An important consideration of the above data deserves some further discussion. In the above papers, we showed that DRC has a negative impact on performance when maskers are close to the target (15°) but not when they were far (90°). We interpreted this result as supporting our assertion that the reduction and fluctuations of ILD due to DRC were not severe enough to impair performance when ILDs were large. However, another contributing factor was likely the presence of ITDs. It may be the case that when the spatial separation was small, ITDs alone were not sufficient to perform the task, requiring listeners to rely on ILD and therefore suffer due to their degradation. However, ITDs for the larger separations allowed for robust direction of attention, allowing subjects to focus on these cues that were unaffected by DRC. With this possibility in mind, it cannot be said that attentional separation by ILD was necessarily robust to DRC at these large separations; indeed, previous research finds a negative effect of DRC even with ILDs as large as 16 dB (Kalluri and Edwards, 2007). Nevertheless, the fact remains that effects of DRC on performance measures in spatial attention tasks may likely not show significant differences at larger angles, and future experiments should take this into consideration.

We also performed one additional experiment after the manuscript show in in Section 4.1 was submitted for publication, and present the results here. This experiment was identical to Experiment 1 as described in the manuscript, except that we used the fast attack time (11 ms) paired with the slow release time (730 ms). By doing this, we allow ourselves to gain further insight into teasing apart the contribution of attack and release time constants in the demonstrated effects of DRC on spatial attention. The results in this experiment were very similar to those in the Fast condition of Experiment 1. Spatial thresholds and absolute performances were very similar across all time constant combinations, and are not shown here. Individual differences revealed a significant negative effect of independent DRC compared to both linear processing and linked compression for spatial thresholds (Wilcoxon sign-rank test; $p<0.01$) and performance with maskers at ±15° (t-test, $p<0.01$ for linear, p<0.05 for linked). These results indicate that a fast attack time, even paired with a slow release time, is sufficient to negatively affect spatial attention.

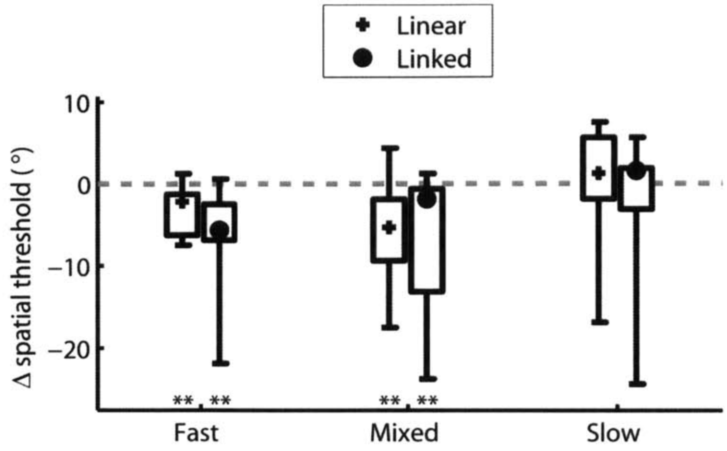

Figure 4.8 – Individual differences in spatial threshold including data from the mixed experiment (fast attack / slow release)

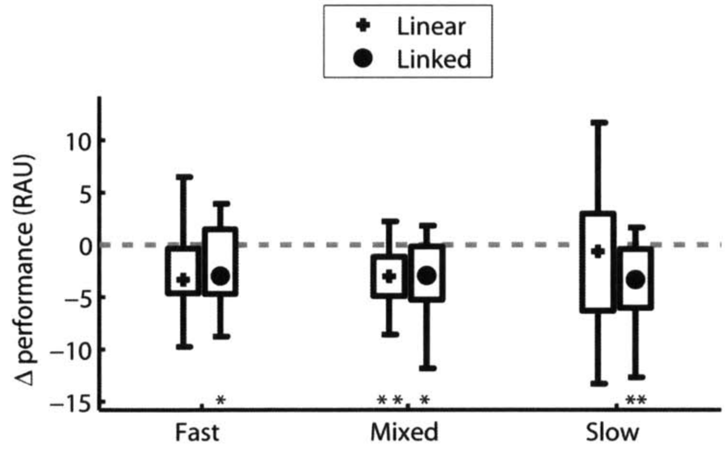

Figure 4.9 – Individual differences in performance when maskers were located at ±15° including data from the mixed experiment (fast attack / slow release)

The manuscripts and additional data presented here add an important piece to the puzzle of spatial perception with DRC. Namely, while we confirmed that DRC has no effect on spatial benefit of maskers separated by 90° (Marrone et al., 2008a), we showed that independent DRC negatively affects the spatial benefit afforded by smaller separations, and that larger spatial separations are required for equivalent performance to linear processing or linked compression. We subsequently demonstrated that this effect was due to the influence of DRC on ILD, rather than on non-spatial features of the experiment stimuli. Taken together, these manuscripts support the conceptual framework we have outlined in Chapter 2, and suggest that strategies such as linked compression may prove to be beneficial for hearing aids. However, we have only tested our hypotheses in normal-hearing subjects. While, intuitively, the same principles outlined in Chapter 2 may apply equally to hearing-impaired as normal-hearing listeners, the next chapter of this thesis is dedicated to demonstrating this idea more concretely with a computational auditory model that includes simulations of sensorineural hearing loss.

Chapter 5

Auditory Model

An important limitation of the behavioral results in Chapter 4 is that they come from normal-hearing listeners, while the primary motivation of this research involves DRC in hearing aids. Hearing-impaired listeners may show a different pattern of results, and this potential difference must be considered as this line of research moves forward. In this chapter, we develop a computational model of the human binaural auditory system that can give us insight into how differences between normal-hearing and hearing-impaired listeners may affect the trends shown in the previous chapter.

To this end, we built a model from existing, published model structures, use this model to analyze the same type of digit mixtures used in our experiments, and compare the pattern of results obtained to those from our behavioral experiments described in Chapter 4. This comparison allows us to build a deeper understanding of the results of our behavioral experiments, and to more concretely expand the conceptual framework outlined in Chapter 2. We also describe modifications to the model designed to represent some basic aspects of hearing loss that may affect our results. In particular, we are interested in modeling the reduced compression that occurs in the impaired inner ear as well as the increased bandwidth of auditory filters (Moore, 2007).

5.1 Model components

Most of the model components used here are based on the model structure described by Faller and Merimaa (2004). Using an existing model structure allows us to focus our investigation on the effect of compression rather than the model development, and Faller and Merimaa have put together a good phenomenological model of auditory processing from the periphery to binaural cue extraction. We must make a few modifications to incorporate hearing loss into the model, and we also added a new model stage that accounts for the fact that ILD resolution is range-dependent (Searle et al., 1976; Koehnke and Durlach, 1989). Finally, we describe a process to quantify the discriminability of sound sources by their detected ILD, and use this metric to model the behavioral performance shown in Chapter 4.

5.1.1 Auditory periphery

A schematic overview of one band of the peripheral model is shown below in Figure 5.1. The first model stage captures the filtering and transduction of sound by the inner ear. Sounds are first passed through a gammatone filter bank (Patterson et al., 1995). In the normal-hearing case, the equivalent rectangular bandwidth (ERB) of the filters was set according to Glasberg and Moore (1990). Code to implement these filters is freely available from Slaney (1998). For most of the results discussed in this thesis, we focus on the 3000 Hz band. ILD varies with frequency, and so other filter bands have different overall magnitudes of ILD, but the results we are interested in, namely, the relative trends with DRC, are very similar across all bands (we also looked at 500 Hz, 1000 Hz, 2000 Hz, and 4000 Hz).

The outputs of these filters are put through a stage of compression and neural transduction as proposed by Bernstein et al. (1999). Specifically, the envelope of the filter output is extracted using a Hilbert transform and compressed by raising it to the power 0.2 (for normal hearing). The fine-structure is then combined with the compressed envelope; computationally, this is done by multiplying the original waveform point-wise by the envelope raised to the power -0.8 (1 is subtracted from the exponent 0.2 to cancel the envelope component of the original waveform). This division can result in significant precision errors in epochs of the waveform with very low energy; in this case, the envelope value is very small, resulting in very large values after being raised to a negative exponent. In practice, we find that this can produce transient artifacts in our speech mixture signals. To avoid these artifacts, we only applied compression to samples where the envelope had a sufficiently high level: specifically, we used a threshold of 0 dB SPL (20 μPa). Due to our use of power-weighting, waveform epochs with envelope values below 0 dB SPL will have negligible effect on our final evaluation of ILD (see Sec. 4.1.2 below). By applying compression only above this threshold, any appreciable signal level would receive compression, and small values that should not have any effect on our model results would not yield these transient artifacts.

Neural transduction is modeled by half-wave rectifying and squaring this result, and passing the output through a fourth-order low-pass filter with a cutoff frequency of 425 Hz to simulate the loss of synchrony of auditory nerve fibers to the fine structure of high frequency stimulation. This low-pass filter effectively computes the time-varying power of the compressed waveform using a time window with roughly 2.4 ms width. It is likely that this operation has little effect on the computation of ILD, which also involves computing the average power over a longer time window (10 ms). However, we opted to keep as many components of our model true to the tested model as possible, as these details are not the focus of this work.

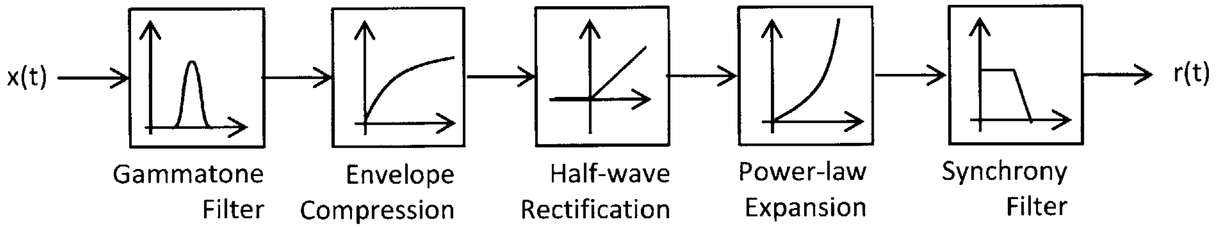

Figure 5.1 – Overview of a single band of the peripheral auditory model

To modify the model for hearing loss, we were interested in capturing two basic phenomena that are typically associated with sensorineural hearing loss and likely to affect the computation of ILD: (1) the loss of dynamic range compression in the inner ear, and (2) the broadening of auditory filters. Both of these changes are captured by the bandwidth multiplier (B) and compression (α) parameters from Moore and Glasberg (2004):

$$B = 10^{0.01 \cdot HL}$$

$$\alpha = \frac{1.7906}{9 - \frac{1}{10} \cdot HL}$$

where HL is the amount of hearing loss expressed in decibels, and the latter equation has been solved for α using equations 4-6 in Moore and Glasberg (2004) and the accompanying boundary conditions. This equation results in values of α that vary between roughly 0.2 for HL = 0 (normal hearing) and 0.6 for HL = 60. The value of B varies from 1 for HL = 0 to roughly 4 for HL = 60. Values of these parameters for increasing levels of hearing loss are plotted below in Figure 5.2.

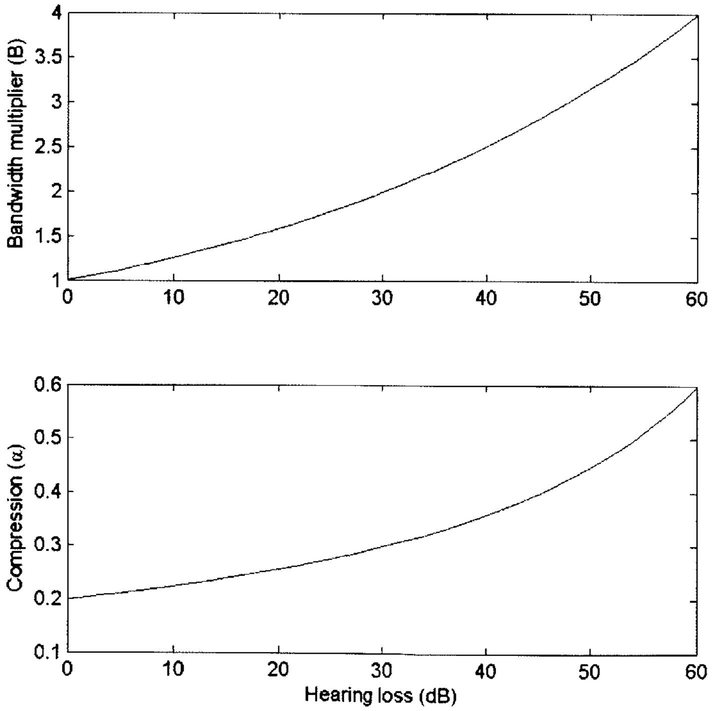

Figure 5.2 – Bandwidth multiplier (B) and compression exponent (α) as a function of amount of hearing loss in dB.

We are not explicitly interested in hearing thresholds with varying degrees of hearing loss. It is assumed that analyses here represent the cues being processed when stimuli are amplified to audible levels. Therefore we are interested in the effects of various model manipulations on ILD and its utility, but not of overall audibility of the stimuli.

5.1.2 Binaural Processing

The outputs of the model described above for both left and right-ear stimuli were then used to compute ILD, following a procedure similar to that described by Faller and Merimaa (2004) and illustrated in Figure 5.3. Specifically, the time-varying power of each waveform was computed using an exponential window with a time constant of 10 ms, and the ratio of power between the right and left signals was computed and converted to decibels. For efficiency, we simplified the Faller and Merimaa model by removing the computation of interaural coherence and ITD as we were not interested in these cues. However, this simplification necessitated a minor change; while Faller and Merimaa computed ILD using left and right windows that were offset to compensate for the estimated ITD, we simply use zero temporal offset. Since the maximum ITD offset used by Faller and Merimaa was 1 ms, which is an order of magnitude below the 10 ms time constant of the time window used for computing power, this offset will have little effect on the ILD output.

To use the output of this model, we need to accurately capture the limited precision with which human listeners can use ILD. Internal noise and limits of the precision of the neural coding of ILD limits the perceptual reliability of small differences in acoustic ILD. Additionally, ILD resolution is dependent on the range of ILDs presented (Searle et al., 1976; Koehnke and Durlach, 1989), and this fact will likely have important consequences when modeling the utility of ILDs reduced by DRC or expanded by hearing loss. We therefore used a context coding model (Durlach and Braida, 1969) that adds two sources of noise to the ILD output to capture both an absolute minimum limit to ILD resolution as well as the range-dependent effects. We fit the variance of these noise sources to roughly match observations published in the literature.

The first noise source is referred to as sensory noise, which is Gaussian with zero mean and a fixed variance β^2. This fixed variance represents the lower limit of ILD discriminability achievable by listeners, as might be revealed by a just-noticeable-difference (JND) experiment paradigm. As described below, we fit this parameter so that the model could reliably distinguish between a 0 and 1 dB ILD.

The second noise source is referred to as memory noise, which was zero-mean Gaussian noise with a variance γ^2, equal to the product of the squares of a constant parameter G and a range parameter R. The range parameter R represents the expected range of ILDs in a given experiment or block of trials. G is a linear multiplier that tells us roughly how many distinct location within the range R can be resolved. For example, a value of G equal to 1/10 indicates that memory noise is roughly one tenth as wide as the full range of ILDs. Put another way, a listener could resolve roughly 10 distinct sound locations before their ILD differences began to overlap significantly. Memory noise captures the fact that localization errors increase when a wider range of ILDs or azimuths are used in an experiment (Searle et al., 1976). To implement these noise sources into our model, we convolved the histograms of observed ILDs for a given simulation trial (see below) with a Gaussian kernel of appropriate width. This procedure is almost identical to adding a time-domain noise waveform to the ILD output, as ILD is integrated over hundreds to thousands of samples to generate each ILD histogram, but is computationally more efficient when relatively many samples make up each time waveform compared to the number of bins used to keep track of ILD histograms.

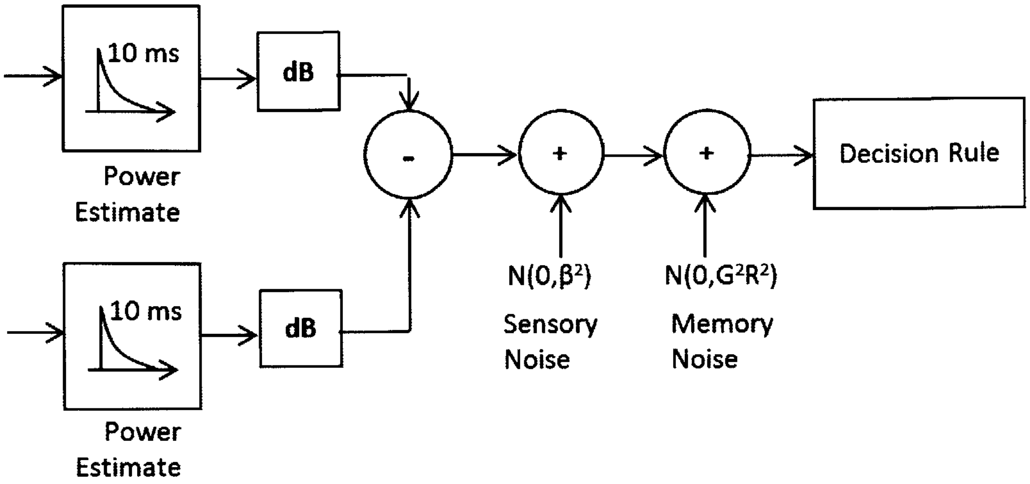

Figure 5.3 – Overview schematic of the binaural model. Inputs are from left and right-ear outputs of a single band of the peripheral model (Figure 5.1)

5.1.3 *Temporal weighting*

In order to use the estimated ILD output to generate decisions in JND and spatial attention simulations, we generated histograms of ILD values that could be used as a representation of a source's estimated location. We weighted the contribution of each ILD sample to its histogram bin by a power-weighting factor that reduced the contribution of low-power regions of the signal and also emphasized signal onsets. Signal onsets frequently dominate ILD and spatial perception (Freyman et al., 1997; Stecker and Hafter, 2002), and this fact may underlie listeners' ability to reliably use the "clean" onset ILD present before a pair of compressors reaches its steady-state attenuation level (Musa-Shufani and Walger, 2006; see Section 5.4). More specifically, in our model, we took the sum of the left and right power signals used to compute ILD, and passed this sum through a feedback adaptation loop (Dau et al., 1996) with a time constant of 20 ms. For an isolated noise burst with a clear onset, this onset weighting will result in roughly the first 20 ms of the stimulus dominating the ILD histogram. However, for acoustic mixtures composed of complex sources, such as speech, the effect of this onset weighting is less obvious. To demonstrate this difference, power-weighting factors computed before and after the adaptation loop are shown below in Figure 5.4.

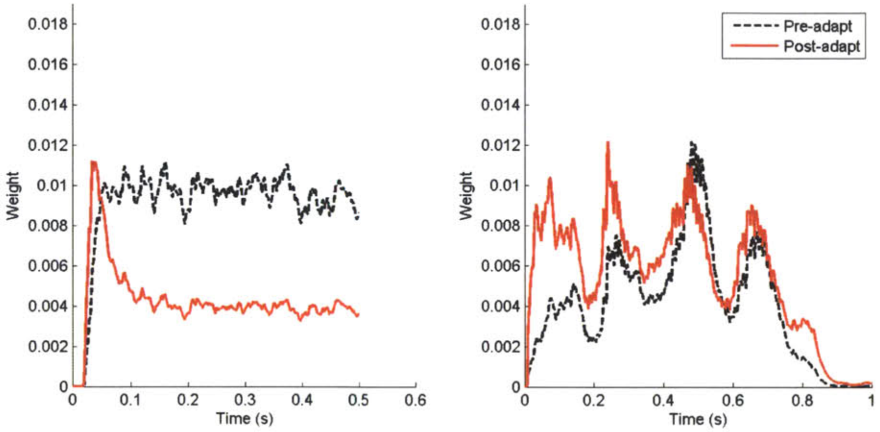

Figure 5.4 – Power-weighting factors for a noise burst (left) and a speech mixture (right). Black dashed line represents the power-weighting factor before the adaptation loop (emphasizing signal onsets) is applied. Red solid lines represent the output of the adaptation loop.

For acoustic mixtures, we generated an ILD histogram for each source in the mixture by using a second weighting factor, called the source-weighting factor, which determined how much each ILD sample contributed to each of the three possible ILD histograms (one target and two maskers). The source-weighting factor for each source indicated when that source was most active or dominant in the mixture. However, in general, multiple sources will be active at any given time. Accurately modeling how the human auditory system parses a mixture into distinct sources is a problem that has gained much attention in the last several decades, but is well beyond the scope of this thesis. Therefore, we used a heuristic rule based on *a-priori* information of the power in each source to determine when a particular source dominates the mixture. Each ILD sample therefore contributed to the ILD histogram associated with whichever source dominated the total power in the acoustic mixture at that time instant. This *a-priori* knowledge can be viewed as a rough model of an ideal set of grouping cues available to human listeners that allow listeners to sort optimally a mixture into sources; cues such as ILD can then be used to focus attention on these grouped objects (e.g., Shinn-Cunningham, 2005; Shinn-Cunningham and Best, 2008). The precise question we are asking by analyzing ILD histograms generated in this way is "how well can ILD predict which source is dominant (according to its power) at any given instant in a given (3 kHz) frequency band?" An example set of these source-weighting signals are plotted in Figure 5.5.

To generate this source-weighting factor for a given source in the mixture, we passed that source (in isolation without the other sources) through the same peripheral auditory model described above. We then computed the average binaural power of this signal using an exponential window with a 20-ms time constant. This time constant matches that used in a previous experiment showing that human listeners could benefit from better-ear glimpses even when those glimpses rapidly fluctuated between left and right ears (Brungart and Iyer, 2012). We computed this binaural power estimate for each source in the mixture and converted to decibels. For our spatial attention simulations, which involved one target and two masker sources, this procedure resulted in three log-power signals. At each time sample, the source-weighting factor for a source was defined as its log-power signal minus the maximum of the other two log-power signals. Therefore, weighting factors above 0 dB indicated that the associated source was dominant in that time epoch. A value near 0 indicated that at least one other source was similar in level at that time epoch, and a value below 0 indicated that another source was dominant. Values below 0 were set to 0 so that no epoch had negative weight for any of the three histograms. Values that were above but still close to 0 (specifically, between 0 and 3 dB) were also set to 0. These small values represent epochs in which another source was similar in level to the desired source, which would result in a corrupted ILD. By setting the weight in these epochs to 0, we discard the contribution of these corrupted ILDs to the source's histogram. Values above 10 dB were set to 10 so that no region had too strong a weight and all regions were limited to finite weights.

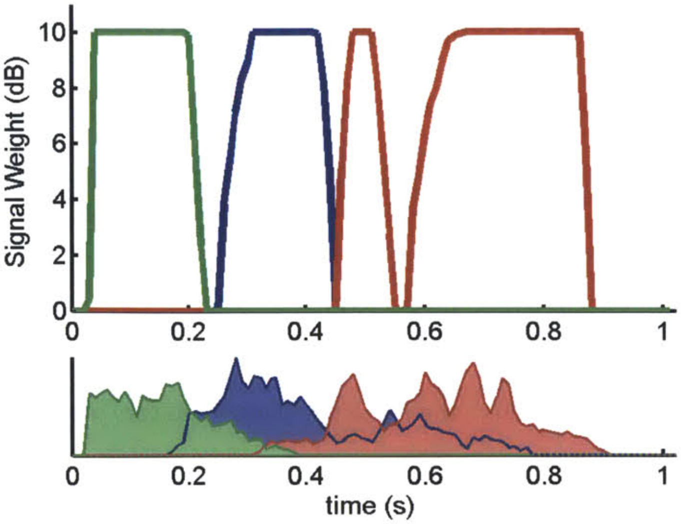

Figure 5.5 – Source-weighting signal (bold lines) for the 3 kHz band output of an example speech mixture consisting of three spoken digits (1 target + 2 maskers). Each source is represented by a color. Each source-weighting signal determines how strongly an ILD sample from each time sample contributes to the corresponding ILD histogram. Here we see good isolation of the regions of the mixture dominated by one source, and rejection of the regions dominated by multiple sources (e.g. at roughly 250 ms and at roughly 550 ms). Shaded areas represent envelopes of the neural response to each isolated source in the mixture. The arbitrary units of the envelope plots are not labeled to avoid cluttering the figure.

5.2 Evaluation of model performance

Our goal with the model is to quantify how well a listener could distinguish between different stimuli based on their ILD. In practice, this is often accomplished by defining a decision rule that would read the ILD at each band to determine which stimulus was most likely to have produced the observed values. For example, if the model were discriminating between two sources from the left and right, and if the ILD at a given instant was 3 dB to the right, the model would decide that the source from the right was likely active.

For our analyses, however, rather than explicitly defining a specific decision rule, we quantified the discriminability of sources based on their ILD by computing a receiver operating characteristic (ROC;

Macmillan et al., 2004). The ROC combines multiple possible decision rules; it can be generated by varying a decision threshold (e.g. between declaring target vs. masker) and plotting for each threshold value the probability of detection (i.e., the probability of declaring "target" given the target was dominant; see Figure 2.1) against the probability of false alarm (i.e., the probability of declaring "target" given that it was not dominant). Discrimination performance is then defined as the area under this ROC curve. This area represents how well ILD can be used to determine the target's presence or absence while making only limited assumptions about the nature of the stimuli or the observer.

For the spatial attention paradigm, three sources (one target and two maskers) were always present. Example histograms can be seen in Figure 5.15. To compute the ROC, we averaged the two masker ILD histograms into a single masker histogram, and computed the ROC by defining a target decision region symmetrically about the median target ILD. We varied this decision region's width to compute the probability of detection and false alarm, and computed the area under the resulting ROC curve.

5.3 Parameter fitting

Most of our model components come from previously published work, and as such, we need not do any extra work to fit the model to human hearing abilities. However, with the addition of the context coding model to account for range-dependent ILD resolution, we have added three model parameters: one parameter, β, representing sensory noise standard deviation, and two parameters, R and G, representing memory noise range and standard deviation per range, respectively. Our goal is not to describe the effects of these parameters on performance, but instead to set realistic values for these parameters to investigate the effects of compression. In this section we describe how we selected values for each of these parameters.

5.3.1 Sensory Noise

To fit the sensory noise parameter β, we ran the model using a two-interval JND paradigm, and adaptively varied β based on the model results. Specifically, using an initial value of β=0 dB, we computed the model performance, as described above, given a white noise stimulus with 0 ILD in one interval and with 1 dB ILD in a second interval. We then adaptively varied the value of β using the simple procedure described below to converge on a performance value (area under the ROC curve) of 71%. These values were chosen to represent commonly used experiment paradigms in the literature, which find ILD JNDs corresponding to 71% or 75% correct performance of around 1 dB for both normal-hearing and hearing-impaired listeners (Hershkowitz and Durlach, 1969; Domnitz and Colburn, 1977; Koehnke et al., 1995; Musa-Shufani et al., 2006). When performance was above 71%, we increased β by the current step size (using an initial step size of 1 dB). When performance was below 71%, we reduced β by the step size and cut the step size in half. We terminated this process once the step size reached 10^{-3} dB. The values of β for varying levels of hearing loss are plotted in Figure 5.6.

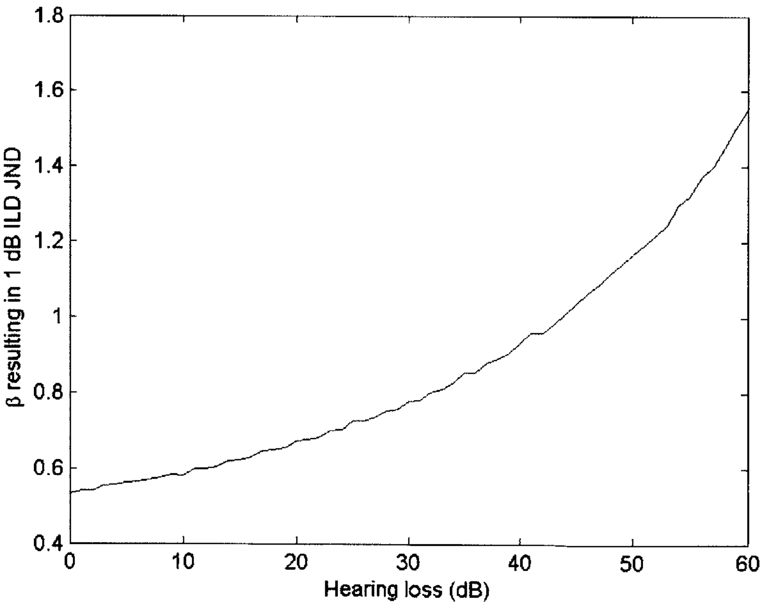

Figure 5.6 – Sensory noise parameter β for varying degrees of hearing loss

Due to the loss of compression with increasing hearing loss (see Figure 5.2, bottom), β increases with the amount of hearing loss. This result comes from the constraint that the ILD JND should be 1 dB regardless of the amount of hearing loss. If β were to be held constant, the loss of compression would result in expanded ILDs, and therefore better performance given the same acoustic ILD. However, this is not observed in the literature (Koehnke and Durlach, 1989; Musa-Shufani et al., 2006); instead, hearing-impaired subjects often have equivalent or worse performance in ILD-related tasks than normal-hearing listeners. By holding JND performance constant across all levels of hearing impairment, we are modeling hearing-impaired listeners who still retain the ability to use ILD as well as their normal-hearing counterparts. It is important to consider that not all hearing-impaired listeners have this ability, and any changes made to hearing-aid compression algorithms aimed at preserving ILD may have reduced or no effect on individuals with severely impaired binaural hearing.

5.3.2 *Memory Noise*

Memory noise variance is a product of the squares of a linear scaling parameter G and a range parameter R. To determine G, we look to the literature. One excellent meta-analysis looked at exactly this question for a similar context-coding model of free-field sound localization in both horizontal and vertical planes (Searle et al., 1976). Analyzing the results of eight prior studies using a combination of ILD, ITD, and monaural level differences as the decision variable, Searle estimated the memory noise parameter G to be 0.062, or roughly 1/16.

Deriving the range R, however, is a tricky issue for our paradigm. Often, R is defined as the difference between the maximum and minimum cues presented in a given experiment (e.g., Shinn-Cunningham, 2000). However, such a procedure will not take the dynamic nature of DRC into account, and will therefore not represent the true range of ILDs encountered in realistic settings. In general, we wish to derive R from the stimuli used in a given paradigm, but it is not clear how to do so in our context in a psychophysically meaningful manner, as demonstrated in the following paragraphs.

Due to the overall reduction of ILD, we might expect DRC to roughly cut the range of ILDs by a factor of the compression ratio (e.g. Figure 5.7A). On the other hand, due to the temporal dynamics of the compressors, a sound's ILD can in fact be "pushed" more laterally at its onset due to a preceding stimulus; DRC can therefore result in a wider absolute range of ILDs than without DRC (e.g. Figure 5.7B and Figure 5.8). However, in this case, we can see that while the total range of ILDs encountered has increased, the centroids of each lobe of the ILD histogram have still been brought closer to zero; the net effect of DRC may still therefore be a small reduction in ILD range, although this reduction is much less pronounced than in the first scenario. Lastly, sounds with levels below the compression threshold will generally have ILD equal to that without DRC, and thus R will not be altered (e.g. Figure 5.7C); in this case, ILD resolution will not change.

Everyday settings likely contain a mixture of all of the above-mentioned types of situations. Intuitively, then, we might still expect bilaterally independent DRC to result in some reduction of the range of ILDs expected by a listener, but to also leave the ILD in some time windows equivalent to, and perhaps in some situations greater than, the ILD that would occur without DRC. There is unfortunately insufficient data in the current literature to derive a simple model of how to adjust or compute R for a given DRC setting. For our purposes, we derived R uniquely for each compression setting and level of hearing loss by repeatedly generating speech-mixture samples (the same that will be used in the attention model simulations; Sec. 4.4) using masker digits located at ±70° (note that the ILD from our HRTFs are largest at approximately 70°). We generated ILD histograms as described in Sec.5.1.3, and defined R as the range of ILDs capturing 80% of the ILD histogram about its median. By using only 80% of the histogram mass, we can exclude small tails of the ILD histogram that may not accurately represent the range of ILDs expected by the subject due to their low prevalence. We can intuitively grasp the effect of this tailexclusion by looking again at Figure 5.7B. Were we to use the full histogram, the compressed stimuli would result in a larger R, and therefore poorer ILD resolution compared to the linear case, and consequently, poorer performance. If subjects are able to adapt to any extent to the reduction of mean ILD by independent DRC, then, using the full histogram as the range will underestimate performance in this condition. However, by limiting R to span 80% of the histogram, we obtain a comparable range of ILDs with DRC (11.6 dB and 11.2 dB for linear and compressed, respectively, in this example).

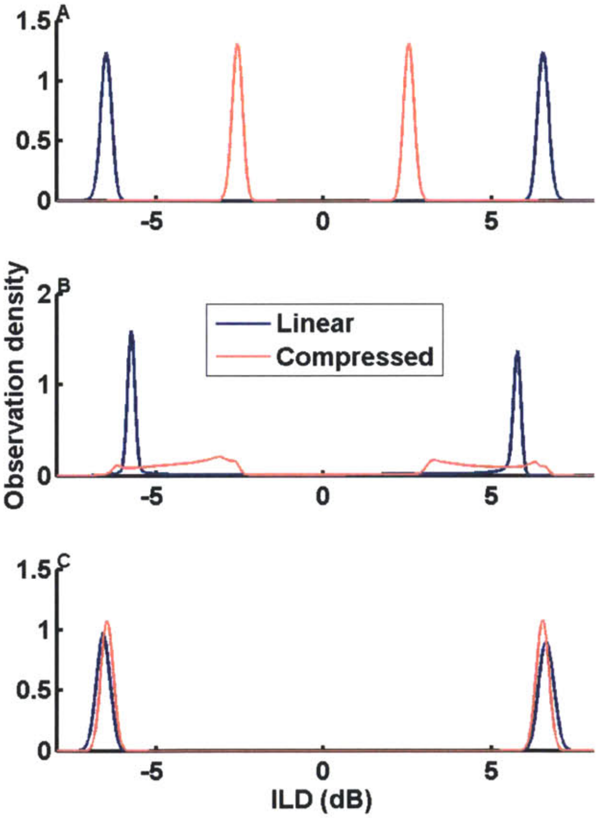

Figure 5.7 – Example ILD histograms in the 3 kHz band, generated for stimuli at ±70° for determining the range parameter R, using linear processing or independent compression. Attack and release time of the compressors were set to 100 ms. A: Two 1-s, 65 dB SPL noise tokens an either +70° or -70°. Each noise token was played and analyzed separately. B: 3 kHz tone at 65 dB SPL, alternating between ±70° every 200 ms. C: Same as A, but the level was set to 35 dB SPL.

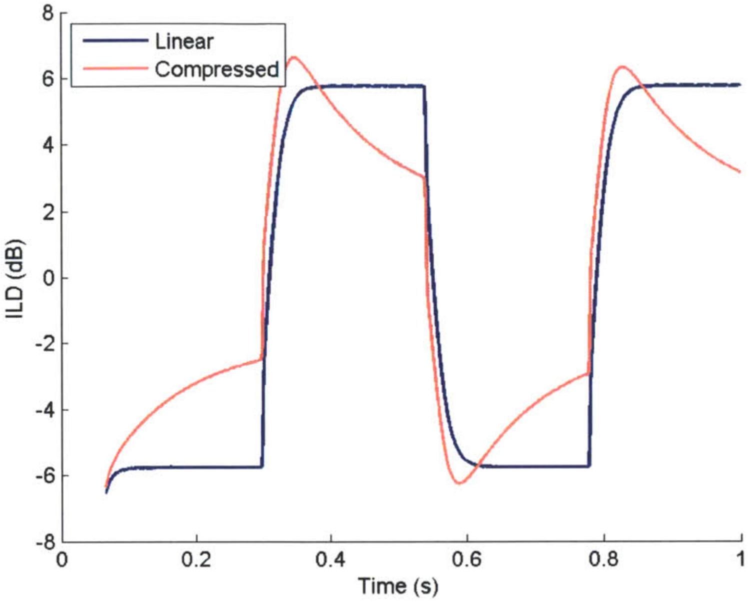

Figure 5.8 – ILD trace for panel B of Figure 5.7, in which a pure tone alternates between ±70°. The ILD at the onset of each alternation is pushed to be more lateral than without compression.

Summed histograms, generated using the response of the 3 kHz band to 100 speech mixture samples, are plotted in Figure 5.9, and the resulting ranges are summarized in Table 5.1. We can see from this table that the independent compression reduces the estimated range of ILDs encountered, but not quite by a factor of the compression ratio (3 in this case). Notice that there is a slight difference between the ILD histograms for linear vs. linked compression, with a corresponding, small increase in range. This difference is due to a difference in the power-weighting signal used to generate the histograms, and is related to why independent compression yields poorer overall SNRs when applied to speech in noise (Wiggins and Seeber, 2013). Portions of the acoustic mixture where multiple sources overlap tend to be higher in level than when only one source is active, and are therefore attenuated more when compression is applied relative to when it is not. Such epochs have ILDs that are some combination of the contributing sources' ILDs. Thus, when compression is linked, the ILD histogram is relatively less dominated by these epochs than the histogram when no compression is applied. In a similar sense to speech in noise, the “effective SNR” is increased, but here, “signal” refers to any of the clean sources rather than just the target, and noise refers to epochs when multiple sources corrupt each other’s ILD. This effect is very modest in our model and likely of little consequence, especially when the source weighting factor is added (eliminating the contribution of these low-SNR epochs).

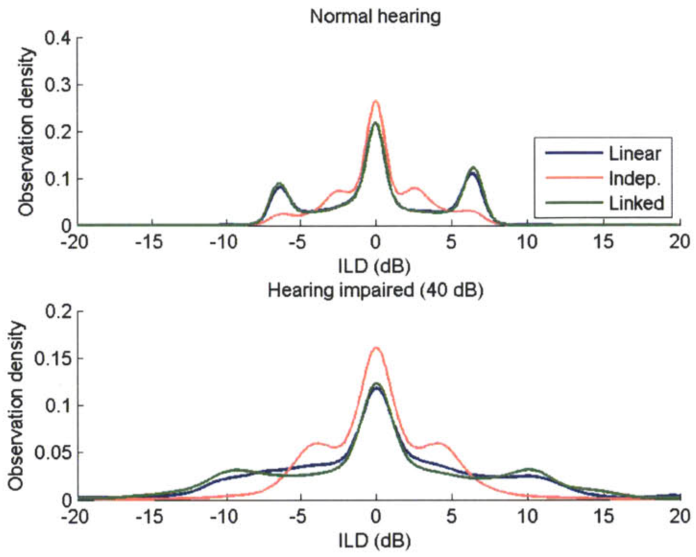

Figure 5.9 – ILD histograms for speech mixtures used to generate the memory noise range parameter, for a normal-hearing model and a model with a flat 40-dB hearing loss.

There is also a notable difference between the normal-hearing and hearing-impaired histograms. Specifically, we see that the peaks corresponding to the digits at ±70° (particularly for the linear and linked compression cases) have been smeared out in the hearing-impaired histograms. This is due to the increased bandwidth of the hearing-impaired peripheral model, resulting in a higher degree of corruption of ILD by the mixture of multiple sound sources (see Section 5.5.1 and Figure 5.17).

	Linear	Independent compression	Linked compression
Normal hearing	12.5	7.8	12.6
Hearing impaired (40 dB)	17.9	10.0	19.9

Table 5.1 – Range of ILD values, in dB, used for the memory noise parameter R for the 3 kHz band. Values represent 90% of the ILD histogram mass about its median, given 100 samples of speech digit triplets in which the maskers were located at ±70°.

5.4 Results: JND paradigm

As a first test of our model, we simulated the experiment of Musa-Shufani et al., (2006), measuring the ILD JND of our model under various compression conditions. Stimuli were 1-s bursts of noise filtered into third-octaves bands. Each burst was given a 20-ms $\cos^2$ on/off ramp and set to 75 dB SPL. Results for noise bands centered at 500 Hz and 4000 Hz were analyzed and qualitatively compared to the behavioral results reported by Musa-Shufani and Walgner. Since we use the data reported by Musa-Shufani et al. to fit the sensory noise parameter β, we expect JNDs for the linear case (ratio of 1:1) to match their data. We are interested in the change in JND due to compression at the higher compression ratios.

To determine JND, the stimuli were presented to the model in two intervals. In the first, the noise was played with zero ILD. The ILD in the second interval varied from 0 to 10 dB, in 1 dB steps. Model performance was computed for each ILD value as described above in Section 5.2, resulting in a data set that described performance as a function of the ILD in the second interval. We fit these data with an inverse normal cumulative distribution function where the mean was constrained to be 0 (so that performance passed through 0.5, or chance, for an ILD of 0). The variance parameter of this function was chosen to be that which minimized the mean squared error of the data compared to the fit. The model JND was then defined as the ILD for which the fitted curve reached a value of 71%. These data are plotted for the normal-hearing model in Figure 5.10. The JNDs values are plotted in Figure 5.12 and Figure 5.13.

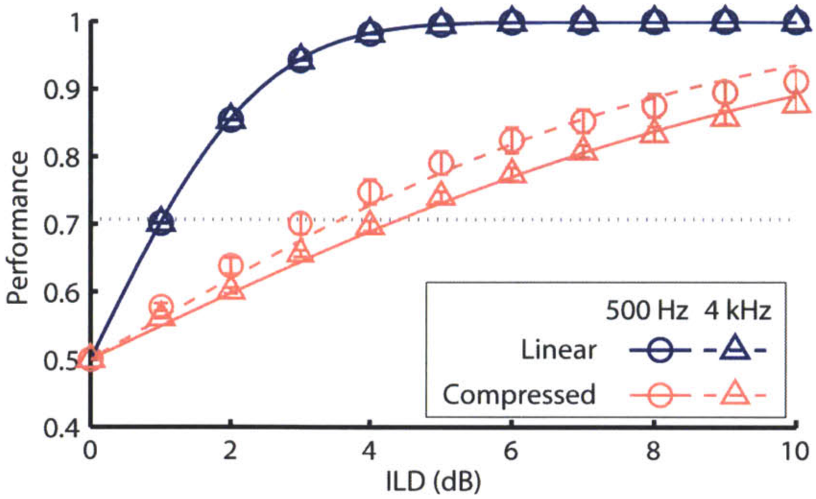

Figure 5.10 – Normal-hearing model performance in a two-interval ILD discrimination task. Symbols represent mean performance values, error bars denote standard deviation of results across 20 repetitions for each point. Solid and dashed curves represent the least mean-squared error fit of an inverse normal cumulative distribution function with mean 0 to the performance data. The grey dotted line represents 71% performance; JND was defined as the ILD resulting in this performance level.

Overall, the model qualitatively captured the effect of increasing compression speed and compression ratio on ILD JND reasonably well, as shown by Figure 5.11 and Figure 5.12. As expected, JNDs were approximately 1 dB for compression ratios of 1:1 (the model was fit so that this would be the case, albeit with wide-band instead of narrow-band noise). There was little to no effect of compression with an attack time of 200 ms, a slight effect for higher compression ratios using an attack time of 20 ms, and a considerable effect using an attack time of 2 ms. The hearing-impaired model performed similarly (Figure 5.13), indicating that the effect of acoustic compression on JND is similar for the normal-hearing and hearing-impaired models. The robustness to moderate and slow compression is due to the onset-emphasis in our power-weighting (Section 5.1.3); when the model was run without this onset-weighting, the compressed ILD dominated and JNDs were much more susceptible to DRC, increasing by almost a factor of the compression ratio (not shown). The behavioral data appears to suggest a possibly larger effect of compression on JND for a 20 ms attack time than suggested by the model. The attack time for which our model begins to show elevated JNDs is related to the time constants used in both the computation of power and the feedback adaptation loop, and this difference, if real, could help refine these stages of our auditory model in future work to better understand the temporal characteristics of binaural processing by human listeners.

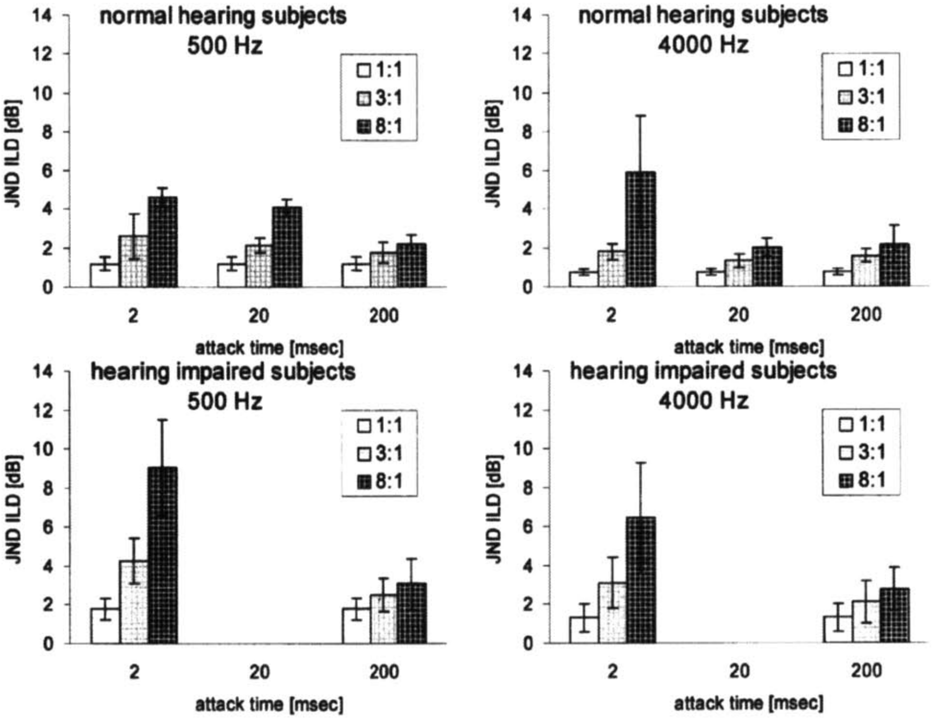

Figure 5.11 – Behavioral data from Musa-Shufani and Walgner, 2006. Mean and 95% confidence intervals are plotted.

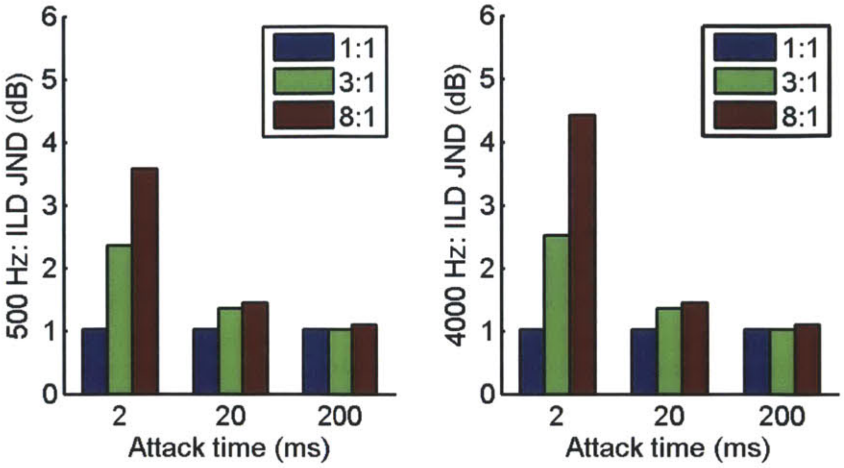

Figure 5.12 – JND results for a normal-hearing model using a 500 Hz band (left) and a 4000 Hz band (right). Different colored bars represent results for compression ratios of 1:1, 3:1, and 8:1.

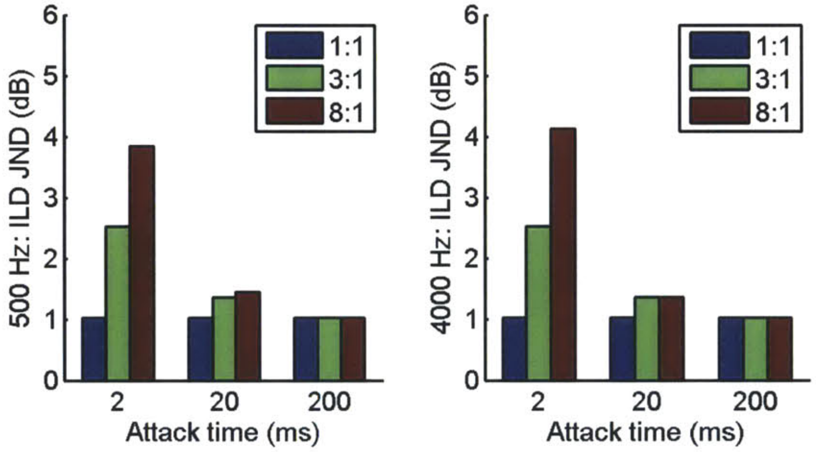

Figure 5.13 – Same as Figure 5.12, but with a flat 40 dB hearing loss in the model

In the behavioral JND data, there was an effect of group (normal-hearing vs. hearing-impaired) on JNDs, but no interaction between group and either attack time or compression ratio. By contrast, our hearing impaired model was fit to have identical ILD JND when no compression was applied. This choice was motivated by the observation that some hearing-impaired subjects have as good ILD sensitivity as their normal-hearing counterparts; however, larger between-subject differences amongst hearing-impaired listeners often result in lower mean performance for the hearing-impaired subject group. In some cases, these individual differences may be due to higher-level processing stages, including cognitive ability, that will not be captured by our peripheral model. In other cases, there may be peripheral deficits resulting in poorer ILD selectivity. We could model the latter case using larger values for the sensory noise parameter β, and as a result the model would predict poorer JNDs. As an example, in Figure 5.14 we show JND results using a hearing-impaired model with a value of beta two times larger than shown in Figure 5.13. JNDs were increased across the board, but the overall effects of compression remain the same. For the remainder of this thesis, we focus our investigation of hearing impairment on a model version with comparable ILD JND to the normal-hearing model, representing the best-case performance.

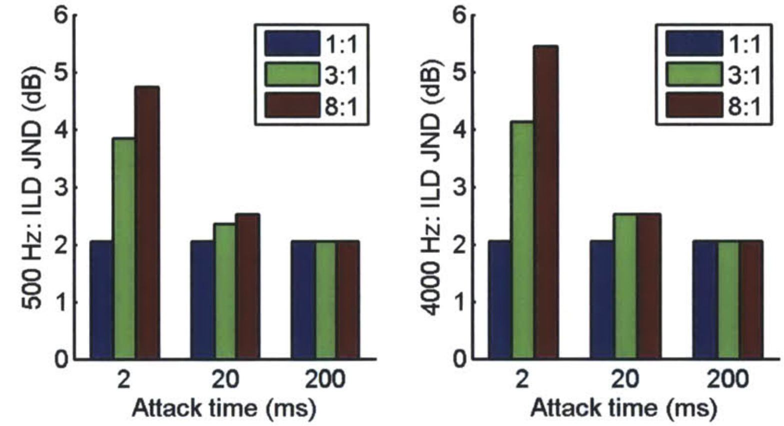

Figure 5.14 – JND results for a hearing impaired model with a larger β (by a factor of 2).

5.5 Results: spatial attention paradigm

The primary goal of our model is to develop understanding of how DRC can affect performance in a spatial attention task, as shown by our behavioral experiments with normal-hearing subjects, and to predict how effects may differ with hearing-impaired listeners. To this end, we replicated the stimuli used in those experiments, and used the model to determine how well ILD could be used to perform the task. Therefore still used the relatively extreme compression ratio of 3:1; future work could investigate how results may vary with specific fitting procedures that match specific cases of hearing loss. However, here we wished to keep as many factors constant as possible, and, as we will see, the underlying principles that affect performance depend more on the total amount of compression, rather than how well that compression is tuned to the hearing loss.

One key difference in the stimuli we used in the model is that instead of creating a sequence of four digits for each target and masker, each stimulus consisted of only one digit from each source. This way, model performance can represent the percentage of digits correctly identified without the need to create a more complex system to parse the longer acoustic mixture into word/digit segments. We systematically varied the azimuth of the masker digits, and for each simulation trial computed the model performance as described in Section 5.2.

All results presented here resulted from the model output at the 3 kHz frequency band. Results from other bands were qualitatively very similar, and are not shown. One notable difference is that at lower frequencies, ILDs are smaller, but because both sensory and memory noise are fitted individually at each frequency band, the absolute ILD magnitude has little effect on ILD discriminability and on the trends with compression, which we are primarily interested in.

5.5.1 Fast compression

Example model outputs using an attack time of 11 ms and a release time of 82 ms (the same compression speed as the "Fast" group in the behavioral experiments in Chapter 4) are shown in Figure 5.15. Linked compression produced a similar result to linear processing, and so this condition is not included in this example figure for visual clarity. The ILD output (panels A and C, bold lines) correctly tracked the position of the active source, and that the source-weighting factor (represented by line thickness and color) correctly identified time regions where one signal is dominating the mixture. The ILD histograms (panels B and D) represent the ILD information after sensory and memory noise have been added. When compression is added (panels C and D), ILD tended to be reduced in some, but not all, epochs of the mixture. Although the model had slightly less memory noise for independent compression (resulting in narrower peaks in panel D vs. panel B), there was still overall more overlap between the ILD histograms in the independent compression condition, resulting in poorer performance.

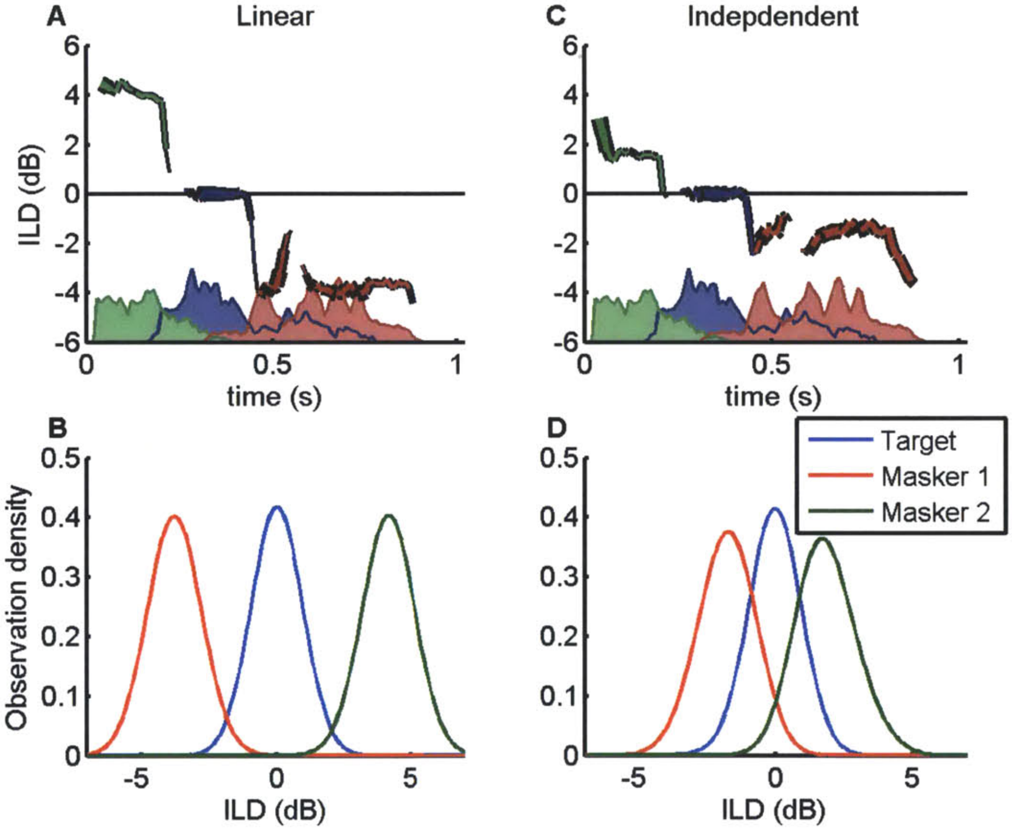

Figure 5.15 – Example model output in the attention simulations for the 3 kHz band with maskers located at ±45°. Figures for linked compression were similar to those for linear processing and so are not included. Top (A and c): Lines represent ILD waveforms. Line thickness indicates the magnitude of the source-weighting factor for the different sources (thicker lines contributed more to the ILD histograms), and the source identities are indicated by color. Provided for easy visual reference, the shaded areas at the bottom of these axes represent envelopes of the neural response to the isolated target and masker signals. Bottom (B and D): ILD histograms generated from the ILD waveforms plotted above.

Results using 1000 simulation trials (each with one target and two masker digits) for each azimuth and compression condition are plotted in Figure 5.16. For linear processing and linked compression, performance quickly rose from near chance (0.5) at low azimuths to near perfect performance above 30°. Consistent with our expectations from the example shown in Figure 5.15 as well as with our behavioral results, average performance was worse when independent DRC was used compared to linear processing, but equivalent for linear processing and linked compression. The model results show relatively large effects of DRC, with performance differences of roughly 15% and 25% at masker azimuths of 15° and 30° respectively. These differences seem larger than the differences shown in our behavioral results, where average performance differences were closer to 5% for maskers at 15°. One possible interpretation of this difference is that our model overestimated the effect of DRC. This could be due, for example, to our method of adapting to a new range of expected ILD being overly simplistic, or our method of segregating the various time epochs of the mixture into source components underperforming the human auditory system (which can use glimpses of short-term better SNR in either ear to improve performance; Brungart and Iyer, 2012). It is not currently known how listeners would adapt to the modified range of ILDs presented with independent DRC; our methods present a reasonable choice for a model, but future work should investigate this specific issue in more detail. Another likely explanation for the seemingly larger effects shown by the model is simply that listeners also had access to ITD, which was not used by our model but helped listeners perform the task. Thus, it is important to keep in mind that these results do not model the behavioral data exactly; rather, they model the trends in the behavioral data that are due to effects on ILD. As such, it is a good tool for making qualitative comparisons, but ill-suited to make quantitative comparisons with behavioral performance data.

The performance difference was most pronounced for moderate masker azimuths, as performance rises rapidly for linear processing and linked compression but more slowly for independent compression. At larger azimuths, performance with independent compression begins to catch up to that of linear processing and linked compression. This result reinforces our idea developed in the previous two chapters that the effect of compression may not be revealed in experiments using only large azimuthal separations between target and masker stimuli, either because ITD difference are more robust at these larger angles or because ILD differences are large enough to be able to withstand corruption by DRC without affecting performance.

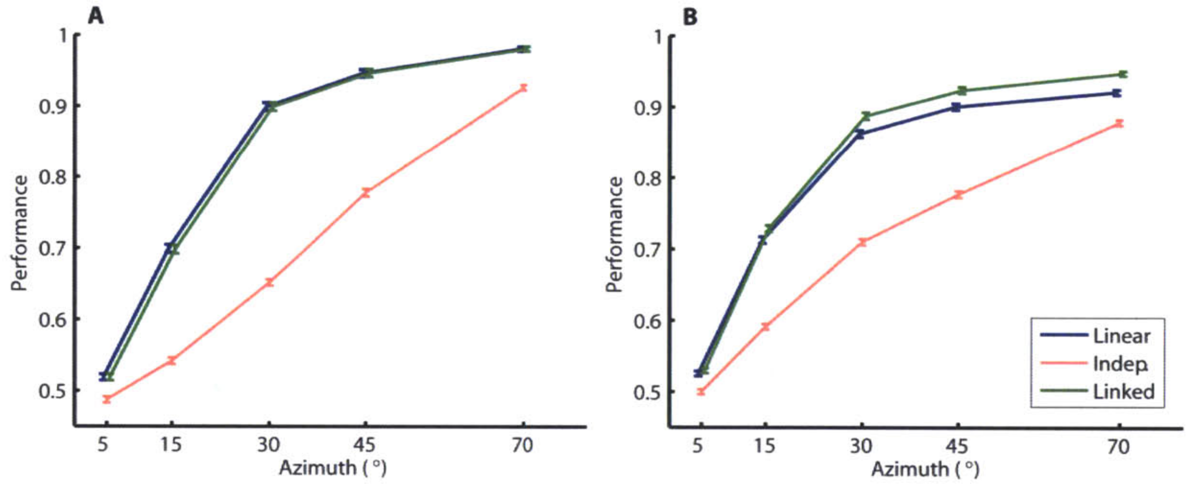

Figure 5.16 – Model results for discriminability of digits by ILD with fast compression using the normal hearing model (A) and the hearing-impaired model (B). Error bars represent the standard error of the mean (each data point represents 1000 trials). Different compression conditions are represented by color, and also plotted with a slight horizontal offset for visual clarity. Without this offset, the linked data falls on top of the linear data as these data points have almost identical performance values.

Performance with the hearing-impaired model is shown in Figure 5.16B. In general, we see that performance was slightly worse overall for the hearing-impaired model than for the normal-hearing model. The reason for this is not immediately obvious; intuitively, we expect the larger ILDs due to the loss of cochlear compression to be roughly balanced by the corresponding increased sensory and memory noise, as demonstrated in the ILD JND simulations. However, by looking into the weighted ILD traces from individual trials, and comparing the normal-hearing model to the hearing-impaired model, we can gain some insight.

This comparison is shown in Figure 5.17 for the linear condition. While ILD values were overall expanded due to loss of compression in the impaired inner ear model (center panel), we can see periods where the ILD of either masker (red and green traces) was brought towards zero, despite the lack of external acoustic compression. This occurred notably around 0.1-0.3 sec and 0.4-0.5 sec, when the target had energy in the mixture in addition to either masker (indicated by the blue dotted line plotted over the red and green shaded areas). It may be that the reduced frequency selectivity of the impaired model resulted in more target energy entering this band, corrupting the ILD of the masker. To confirm this suspicion, we ran the same stimulus again with a hearing-impaired model where we retain the same bandwidth as the normal-hearing model, plotted in the right panel of Figure 5.17. Consistent with our suspicion, the ILD trace returned to roughly match that of the normal-hearing model (left panel), with an expanded ILD range due to the difference in inner-ear compressoin. Over multiple stimuli, we see in Figure 5.18 that overall performance for the hearing-impaired model matches the performance of the normal-hearing model if we set the bandwidth in this manner. In either case, however, DRC had a larger effect on performance than the effect of hearing loss and bandwidth.

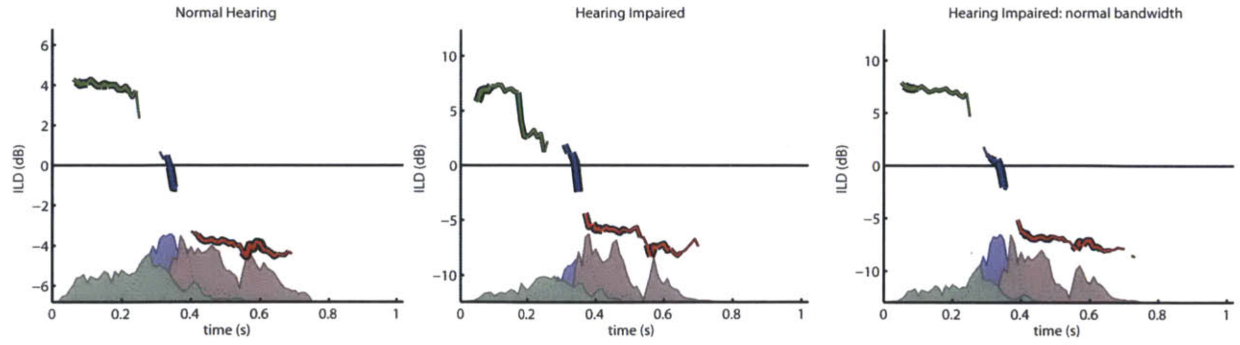

Figure 5.17 – Comparison, for the same stimuli, of the weighted ILD traces from the normal-hearing (left), hearing-impaired (center), and hearing-impaired with normal-hearing filter bandwidth (right). Results are for linear processing (no DRC).

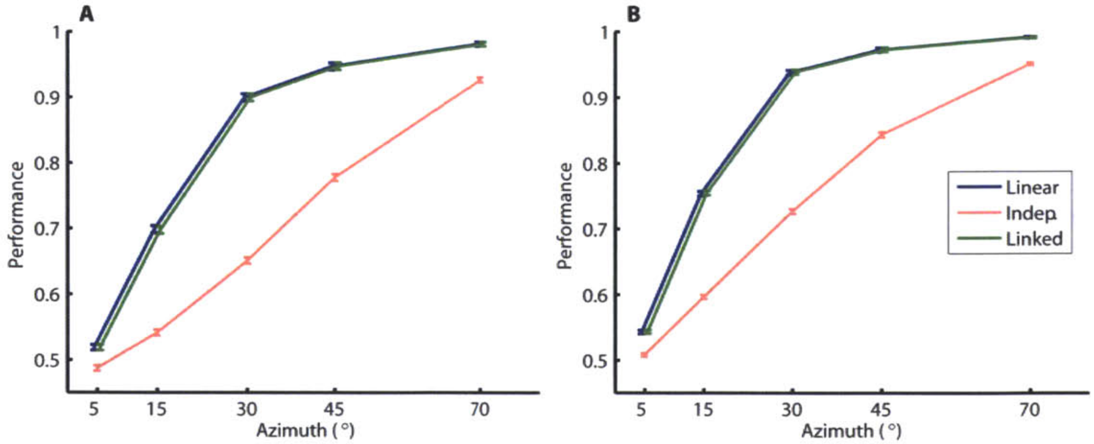

Figure 5.18 – Spatial attention simulation results using the normal-hearing model (A; same as Figure 5.16) and the modified hearing-impaired model (B). The impaired model had inner-ear compression equivalent to a 40-dB hearing loss, but normal-hearing bandwidth for the auditory filters.

Overall, these model results support the idea that the same principles by which DRC impairs normal hearing listeners' abilities to use ILD to attend to speech embedded in a mixture also operate in hearing-impaired listeners. Regardless of hearing loss, independent DRC reduced ILD by as much as the compression ratio. Yet, due to the dynamic nature of the effect of DRC on ILD, the total range of ILD was reduced by less than would be expected from the compression ratio alone. As a result, the reduction in ILDs was not matched with an equivalent reduction in memory noise, resulting in poorer performance. Additionally, when stimuli's ILD differ by an amount comparable to the sensory noise, the reduction of those ILDs by DRC resulted in impaired ability to detect that difference. As a result of these two effects, DRC is likely to impair any subjects' ability to use ILD to discriminate between nearby sound sources in a mixture regardless of hearing loss.

5.5.2 Slow compression

We also ran the model using the slower attack and release time constants used in the behavioral experiments (48 and 730 ms, respectively). All other conditions were identical to those described above.

The results of these simulation runs are shown below in Figure 5.19, for the normal hearing and hearing-impaired models. Unlike the results of our behavioral experiments, we still see a clear effect of DRC in these data. However, the effect here was smaller than with the faster time constants.

While DRC still adversely affected model performance, albeit modestly, there are several reasons consistent with the reasoning presented in Section 5.5.1 why this might not be observed in our behavioral data. Our model results represent discriminability of sources by ILD, not digit recognition performance. Cues such as ITD and spectrotemporal continuity can also contribute to performance and diminish the importance of small differences based on ILD alone. Random lapses of attention might also reduce performance, further obscuring the difference between the conditions shown here. Lastly, random, within-subject variability may eclipse the small effect, making it difficult to observe. The trend when adding hearing loss is similar to that observed when using fast time constants. Overall performance dropped, but the difference between linear processing, independent DRC, and linked DRC was not largely affected.

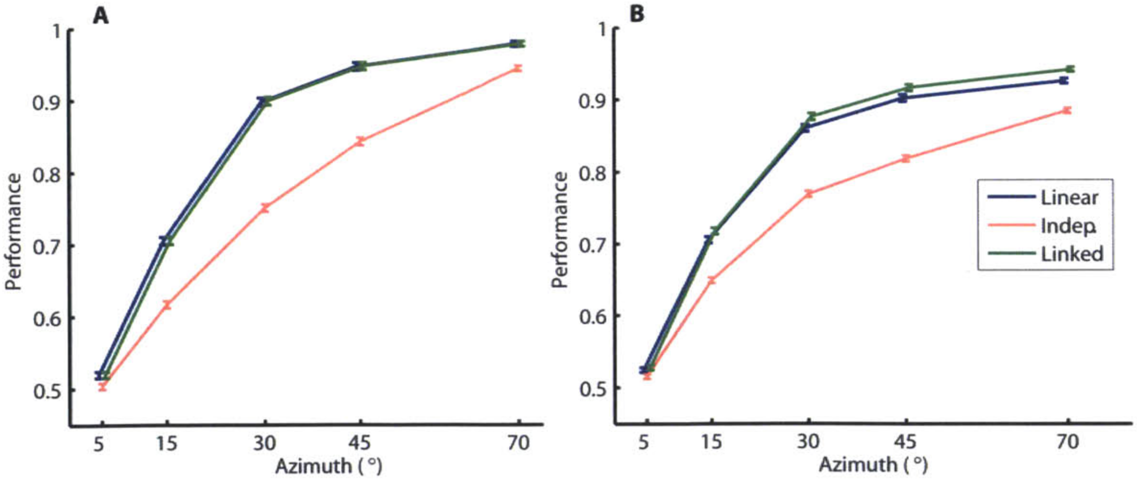

Figure 5.19 – Results of 1000 simulation runs using slow time constants (48 and 730 ms for attack and release time, respectively) for the normal-hearing (A) and hearing-impaired (B) models.

5.6 General discussion and summary

Overall, the model captured the qualitative trends across different spatial perception experiments rather well. ILD JNDs were robust to moderate compression, but increasingly impaired for faster and more extreme compression. Using JND and localization data from the literature to estimate the sensory and memory noise variance in our context coding model produced a pattern of a sharp rise in spatial attention performance with azimuth up to about 30° (for linear processing), and leveling off at higher azimuths. This pattern is roughly consistent with performance in our behavioral experiments as well as more detailed reports of spatial tuning across azimuth in the literature (Marrone et al., 2008b).

It must be kept in mind that our results represent only discriminability by ILD, and do not include effects of ITD or other selection cues such as pitch and timbre. Therefore, all results reported here must be carefully interpreted to demonstrate the changes in ILD utility for similar tasks, but not to quantitatively predict human listener performance in any particular task. There are many situations in which the corruption of ILD may be important even though DRC has mostly no effect on ITD. For example, hearing-impaired listeners often have impaired sensitivity to timing cues (Moore, 2008), which may cause listeners to heavily rely on ILD cues to attend to a given spatial location. The presence of low-frequency noise in an acoustic scene may mask more reliable ITD cues at lower frequencies, again resulting in dependence on ILD, which is more reliable at high frequencies. ILD may also be particularly important for sequentially grouping specific speech sounds, such as fricatives, that have primarily high-frequency content. ITD and ILD are also affected differently by reverberation, and many reverberant environments may render ITD a less useful cue than ILD for spatial perception (Rakerd and Hartmann, 2010; Ihlefeld and Shinn-Cunningham, 2011).

Qualitative trends from the model results were broadly consistent with the observations from our experiments. Independent DRC reduced performance compared to linear processing and linked DRC.

The effect was largest for moderate masker azimuths but virtually disappeared for larger azimuths. The effect of independent DRC was reduced, although not eliminated, for slow compression compared to fast compression.

One weakness of the behavioral experiments is that we were unable to provide a realistic period of acclimatization to the compression ratio to our listeners. While we blocked the behavioral experiment to allow for some rapid adaptation to each compression condition, it is still possible that performance with compressive amplification devices worn continuously for months would improve to the point of showing no difference compared to the linear case. However, insights from our model show why this may not be the case. By appropriately fitting memory noise to each condition, we are roughly modeling human listeners' ability to adapt to modified ranges of interaural cues in the long-term. However, in our model, even though memory noise range was reduced using independent compression, it was reduced by less than the compression ratio due to the dynamic nature and level-dependence of DRC. Nevertheless, at certain times (in particular when the input level was highest and thus the ILD most dominant in our model), the instantaneous ILD of a masker is compressed by as much as the compression ratio, thereby producing poorer segregability of the sound sources. While our method of determining ILD range was largely arbitrary rather than based on specific physiological principles, it represents a reasonable initial model of how subjects may adapt to new ranges of ILD, and this contrast between instantaneous ILD reduction and overall ILD range reduction demonstrates why it may be incorrect to assume that human listeners can simply adapt their spatial perception to an arbitrarily defined compression scheme.

Another idea suggested by our model is that the broadened auditory filters can result in poorer representation of ILD in acoustic mixtures (e.g., Figure 5.17). Of course, the idea of broadened auditory filter impairing auditory grouping is not new. Broadened auditory filters can result in impaired frequency selectivity and ability to form perceptual streams (Gaudrain et al., 2007; Hopkins and Moore, 2011). This result is often attributed to poorer ability to separate the spectrotemporal content into auditory streams. Our results highlight that selection cues, such as ILD, can also be corrupted; broadened auditory filters may impair auditory streaming not only due to an impaired ability to form perceptual streams, but also by corrupting the selective attention cues that allow listeners to select among them. This idea is consistent with the fact that in spatial attention tasks, hearing-impaired listeners enjoy less benefit of spatial separation than normal-hearing listeners (Best et al., 2011).

Our model successfully reproduced the effects of DRC on ILD JNDs and spatial attention tasks in a manner consistent with the conceptual framework outlines in Chapter 2. One difference between our model results and the general predictions in Chapter 2 is in the predicted increased width of ILD distributions. In Chapter 2, we predicted that ILD fluctuations might generally yield an increased width of ILD distributions associated with the target and maskers alike; however, using our model, we generally see that the ILD histograms of the central target was not largely affected by compression Figure 5.15D). The ILD histograms of the lateral maskers can be smeared quite severely (as in Figure 5.7B), but for speech mixtures generally tended be of comparable width, with perhaps some tail toward the original ILD (Figure 5.15D). However, given that memory noise was reduced with independent DRC (Table 5.1), the comparable width of the ILD distributions suggests that the dynamic fluctuations in ILD were an important factor limiting performance. This idea is consistent with the report that a static reduction in ILD did not affect spatial perception as much as DRC, and that DRC increased the perception of image diffuseness (Wiggins and Seeber, 2012). Furthermore, were ILD reduced statically, memory noise would have been reduced by a larger factor (see Figure 5.9), likely resulting in improved performance. Thus while ILD fluctuations caused by DRC did not increase ILD histogram width quite like expected, these dynamic fluctuations were still a critical factor in our results.

Chapter 6

Conclusions

6.1 Thesis overview

In this thesis, we have developed an understanding of how DRC can affect ILD, from simple, isolated stimuli to more complex mixtures. We then presented a review of the existing findings on the effects of DRC on various aspects spatial hearing. Many of these findings seemed intuitively at odds with each other, but by careful consideration of the effects at play we were able to see how these results can be interpreted consistently with each other.

In essence, DRC blurs the spatial representation of sounds by reducing ILD and introducing temporal fluctuations that depend on the interaction between different sound sources at different times in the acoustic scene. This blurring might be perceived as increased image diffuseness with DRC, even when only a single sound source is present (e.g., Wiggins and Seeber, 2012). As a result, a listener's ability to attend to a target speaker in a mixture based on its spatial location can suffer when DRC is bilaterally independent. This blurring is only detrimental, however, when the spatial separation between this target speaker and other interfering sources is sufficiently small such that the spatial images begin to overlap. When sources are far enough apart, even the blurred spatial images are sufficient to distinguish between them.

We tested our ideas by designing an experiment intended to maximize the effect of DRC on spatial perception while keeping intact normally-occurring cues such as ITD. We found an adverse effect of DRC on listeners' abilities to attend to a sound source based on its location; however, this effect, while robust, was modest: performance when maskers were located at ±15° was only reduced by an average of less than 10 percentage points, and spatial threshold increased by typically 5°. This modest effect may be related to our choice to keep maskers symmetrical, limiting the bilaterally-independent effects of DRC for smaller masker azimuths. On the other hand, performance is likely to improve when other differences, including individual speaker characteristics such as pitch, rate, and intensity, are present, as in most realistic settings.

Our auditory model was able to qualitatively account for the trends observed in our behavioral experiment, although it overstated the effect of DRC on behavioral performance. There were no fundamental differences in the effects of DRC between the normal-hearing model and the hearing-impaired model, suggesting that similar potentially problematic effects may occur with DRC in patients' hearing aids. Further research is needed to evaluate this possibility.

6.2 Future work

The work presented in this thesis highlights the need for more thorough research in the area of DRC and spatial perception, particularly for hearing-impaired individuals. In particular, such efforts must acknowledge the distinction between localization and spatial attention, and, along these lines, must address the potential for effects that may only occur at smaller angles of separation than used by many previous studies, but that are nonetheless important for daily life. It may also be important to consider how DRC might affect the effort required to attend to a talker in a mixture, even if direct performance measures in laboratory experiment show only small effects. It is possible that corrupting ILD results in a higher degree of cognitive effort required to attend to a desired source, and this possibility may be meaningful to hearing-aid wearers.

Despite our findings related to spatial attention, it must be acknowledged that DRC has proven to be highly beneficial for hearing aids. Thus while we would not advocate for the removal of DRC from hearing aid processing strategies, we have shown that linking the left and right compressors can improve spatial perception, consistent with previous research (Wiggins and Seeber, 2013). Future research could focus more on the differences between linked and independent DRC related to spatial and other aspects of hearing, aimed at helping to determine if linking two such devices is beneficial enough to warrant the increased design complexity and power consumption requirements that would be required. The effects we showed in Chapter 4 were relatively modest in terms of total performance, and this may point to the need to prioritize other factors such as device battery life.

Our experiments were designed to represent a relatively realistic acoustic scenario, but are still only one viewpoint from which to judge the importance of independent vs. linked DRC on hearing in complex environments. For example, we have discussed how asymmetrical spatial configurations may result in more extreme effects on performance. On the other hand, the presence of other cues such as pitch and voice quality may result in reduced effects. Additionally, objective and subjective effects are likely to vary with larger vocabularies than the digits task used here, and with running speech. Such settings are likely to be more taxing on cognitive resources, which may make it more important to preserve as many attentional cues as possible. On the other hand, the limiting factor in such settings may be the ability to group sounds into sound objects, which is a necessary prerequisite to selective attention (Shinn-Cunningham and Best, 2008). While it is important for future work to develop a more complete characterization of the effects of DRC in these various settings, ultimately, it is each patient's experience that determines the value of any particular hearing aid device. Thus it is also important for future work to consider subjective evaluation by patients using linked or independent DRC, both independent of and in combination with co-varying factors such as device cost and battery life.

An important factor to future research that we have omitted in our discussions is how to approach bilateral compression with asymmetric hearing loss. In these cases, the left and right ears may require different amounts of gain and different compression ratios to compensate for the differing levels of hearing loss in each ear. The amount of amplification required for a sound of a given intensity to achieve the same loudness at each ear may not match the binaural amplification required to produce a centered localization percept (Florentine, 1976; Durlach et al., 1981; Mencher and Davis, 2006). Furthermore, different compression ratios at either ear will greatly complicate the interaction of DRC with ILD. In this thesis we have restricted our analysis to the relatively simple case of symmetrical hearing loss; but further research should investigate how asymmetries may add depth to the issues of DRC and spatial hearing.

We have also shown that slower compression speeds can mitigate the adverse effects of independent DRC. This observation is consistent with previous reports of enhanced benefit of spatial separation using slow compression speeds with extended bandwidth of amplification (Moore et al., 2010). However, many other factors that are not directly related to spatial hearing influence the choice of slow vs. fast compression speeds, and many of these effects are still being debated in the scientific and clinical communities. This research adds to that of Moore et al. to suggest that slower compression speeds may be beneficial when spatial cues are useful, particularly if linking the compressors is not an option.

Even though we attempted to represent several realistic factors of everyday communication, such as having temporally asynchronous speech sources and full-cue HRTFs, our simulations still only covered a specific set of stimuli and conditions. The ultimate test of the effect of DRC must be done in truly real-world settings, where stimuli will be even more varied, other grouping cues will be available, and listeners can learn to adapt their listening strategy in any way that they find beneficial. However, quantitatively measure performance and cognitive effort in such situations remains a challenege.

Consistent with much previous research, our auditory model also demonstrated that poorer representation of ILD under conditions of hearing impairment can result from poorer frequency selectivity in the inner ear. Strategies for enhancing spectral contrast are not a new idea, but here we suggest that this idea might be taken further to enhance ILD in a way as to limit the interference of differing ILDs in different frequency regions. For example, it may be beneficial to compute the ILD at peaks in a stimulus' spectrum, and impose this ILD on a spectral region surrounding this peak to yield sharper ILD distributions associated with a given sound source.

Finally, it will always be important to recognize individual differences in hearing impairment and its effects. As future research on this topic develops, even if linked vs. independent compression is confirmed to be beneficial on average to hearing-impaired individuals, this does not imply that it is beneficial to all individuals. For example, older hearing-impaired individuals who display insensitivity to sound width based on interaural coherence may not benefit from more precise ILD cues. Conversely, it may be important to allow hearing-impaired children access to preserved or enhanced ILD cues so that their auditory systems can learn to use these cues successfully. There are many avenues of research to pursue in this area, and it is our hope that this thesis has contributed one small step toward a better understanding of many of these issues.

IV. References

Ahlstrom, J. B., Horwitz, A. R., and Dubno, J. R. (**2009**). "Spatial benefit of bilateral hearing aids," Ear Hear. **30**, 203–18.

Akeroyd, M. A. (**2004**). "The across frequency independence of equalization of interaural time delay in the equalization-cancellation model of binaural unmasking," J. Acoust. Soc. Am. **116**, 1135–48.

ANSI (**2003**). *Specification of Hearing Aid Characteristics (ANSI S3.22-2003)* (American National Standard Institute, New York).

Arbogast, T. L., Mason, C. R., and Kidd, G. (**2005**). "The effect of spatial separation on informational masking of speech in normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **117**, 2169–2180.

Bauer, R. W. (**1966**). "Noise Localization after Unilateral Attenuation," J. Acoust. Soc. Am. **40**, 441.

Bernstein, L. R., van de Par, S., and Trahiotis, C. (**1999**). "The normalized interaural correlation: accounting for NoS pi thresholds obtained with Gaussian and 'low-noise' masking noise," J. Acoust. Soc. Am. **106**, 870–6.

Best, V., Gallun, F. J., Mason, C. R., Kidd, G., and Shinn-Cunningham, B. G. (**2010**). "The impact of noise and hearing loss on the processing of simultaneous sentences," Ear Hear. **31**, 213–20.

Best, V., Mason, C. R., and Kidd, G. (**2011**). "Spatial release from masking in normally hearing and hearing-impaired listeners as a function of the temporal overlap of competing talkers," J. Acoust. Soc. Am. **129**, 1616–1625.

Van den Bogaert, T., Klasen, T. J., Moonen, M., Van Deun, L., and Wouters, J. (**2006**). "Horizontal localization with bilateral hearing aids: without is better than with," J. Acoust. Soc. Am. **119**, 515–26.

Bolia, R. S., Nelson, W. T., Ericson, M. A., and Simpson, B. D. (**2000**). "A speech corpus for multitalker communications research," J. Acoust. Soc. Am. **107**, 1065–6.

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

Bridges, J. F. P., Lataille, A. T., Buttorff, C., White, S., and Niparko, J. K. (**2012**). "Consumer preferences for hearing aid attributes: a comparison of rating and conjoint analysis methods," Trends Amplif. **16**, 40–8.

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

Bronkhorst, A. W., and Plomp, R. (**1989**). "Binaural speech intelligibility in noise for hearing-impaired listeners," J. Acoust. Soc. Am. **86**, 1374–83.

Brungart, D. S. (**2001**). "Informational and energetic masking effects in the perception of two simultaneous talkers," J. Acoust. Soc. Am. **109**, 1101–1109.

Brungart, D. S., and Iyer, N. (**2012**). "Better-ear glimpsing efficiency with symmetrically-placed interfering talkers," J. Acoust. Soc. Am. **132**, 2545–56.

Byrne, D., and Noble, W. (**1998**). "Optimizing Sound Localization with Hearing Aids," Trends Amplif. **3**, 51–73.

Carlyon, R. P. (**2004**). "How the brain separates sounds," Trends Cog. Sci. **8**, 465–71.

Cherry, E. C. (**1953**). "Some experiments on the recognition of speech, with one and with two ears," J. Acoust. Soc. Am. **25**, 975.

Culling, J., and Colburn, H. (**2000**). "Binaural sluggishness in the perception of tone sequences and speech in noise," J. Acoust. Soc. Am. **107**, 517–527.

Culling, J. F. (**2007**). "Evidence specifically favoring the equalization-cancellation theory of binaural unmasking," J. Acoust. Soc. Am. **122**, 2803–13.

Dahmen, J. C., Keating, P., Nodal, F. R., Schulz, A. L., and King, A. J. (**2010**). "Adaptation to Stimulus Statistics in the Perception and Neural Representation of Auditory Space," Neuron **66**, 937–948.

Dau, T., Püschel, D., and Kohlrausch, A. (**1996**). "A quantitative model of the 'effective' signal processing in the auditory system I Model structure," J. Acoust. Soc. Am. **99**, 3615–22.

Domnitz, R., and Colburn, H. (**1977**). "Lateral position and interaural discrimination," J. Acoust. Soc. Am. **61**, 1586.

Durlach, N. I. (**1963**). "Equalization and Cancellation Theory of Binaural Masking-Level Differences," J. Acoust. Soc. Am. **35**, 1206–1218.

Durlach, N. I., and Braida, L. D. (**1969**). "Intensity perception I Preliminary theory of intensity resolution," J. Acoust. Soc. Am. **46**, 372–83.

Durlach, N. I., Thompson, C. L., and Colburn, H. S. (**1981**). "Binaural interaction of impaired listeners A review of past research," Audiology : official organ of the International Society of Audiology **20**, 181–211.

Edwards, B. (**2007**). "The future of hearing aid technology," Trends Amplif. **11**, 31–46.

Elhilali, M., Chi, T., and Shamma, S. A. (**2003**). "A spectro-temporal modulation index (STMI) for assessment of speech intelligibility," Speech Communication **41**, 331–348.

Faller, C., and Merimaa, J. (**2004**). "Source localization in complex listening situations: selection of binaural cues based on interaural coherence," J. Acoust. Soc. Am. **116**, 3075.

Florentine, M. (**1976**). "Relation between lateralization and loudness in asymmetrical hearing losses," Ear Hear. **1**, 243.

Florentine, M., Reed, C. M., Rabinowitz, W. M., Braida, L. D., and Durlach, N. I. (**1993**). "Intensity perception XIV Intensity discrimination in listeners with sensorineural hearing lossa)," J. Acoust. Soc. Am. **94**, 2575.

Francart, T., Lenssen, A., and Wouters, J. (**2011**). "Enhancement of interaural level differences improves sound localization in bimodal hearing," J. Acoust. Soc. Am. **130**, 2817.

Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (**2004**). "Effect of number of masking talkers and auditory priming on informational masking in speech recognition," J. Acoust. Soc. Am. **115**, 2246–2256.

Freyman, R. L., Zurek, P. M., Balakrishnan, U., and Chiang, Y. C. (**1997**). "Onset dominance in lateralization," J. Acoust. Soc. Am. **101**, 1649–59.

Gallun, F. J., Durlach, N. I., Colburn, H. S., Shinn-Cunningham, B. G., Best, V., Mason, C. R., and Kidd, G. (**2008**). "The extent to which a position-based explanation accounts for binaural release from informational masking," J. Acoust. Soc. Am. **124**, 439–49.

Gallun, F. J., Mason, C. R., and Kidd, G. (**2007**). "Task-dependent costs in processing two simultaneous auditory stimuli," Perception & psychophysics **69**, 757–71.

Gardner, B., and Martin, K. (**1994**). "HRTF measurements of a kemar dummy-head microphone,". Retrieved May 16, 2012, from http://sound.media.mit.edu/resources/KEMAR.html

Gatehouse, S., and Noble, W. (**2004**). "The Speech, Spatial and Qualities of Hearing Scale (SSQ)," Int. J. Audiol. **43**, 85–99.

Gaudrain, E., Grimault, N., Healy, E. W., and Béra, J.-C. (**2007**). "Effect of spectral smearing on the perceptual segregation of vowel sequences," Hearing research **231**, 32–41.

Glasberg, B. R., and Moore, B. C. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–38.

Grantham, D. W., and Wightman, F. L. (**1978**). "Detectability of varying interaural temporal differences," J. Acoust. Soc. Am. **63**, 511–523.

Helfer, K. S., and Freyman, R. L. (**2008**). "Aging and speech-on-speech masking," Ear Hear. **29**, 87–98.

Hershkowitz, R. M., and Durlach, N. I. (**1969**). "Interaural time and amplitude jnds for a 500-Hz tone," J. Acoust. Soc. Am. **46**, 1464–7.

Hopkins, K., and Moore, B. C. J. (**2011**). "The effects of age and cochlear hearing loss on temporal fine structure sensitivity, frequency selectivity, and speech reception in noise," J. Acoust. Soc. Am. **130**, 334–49.

Ihlefeld, A., and Shinn-Cunningham, B. G. (**2008**). "Spatial release from energetic and informational masking in a selective speech identification task," J. Acoust. Soc. Am. **123**, 4369–4379.

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

Jahnke, J. C. (**1965**). "Primacy and recency effects in serial-position curves of immediate recall," J. Exp. Psychol. **70**, 130–2.

Jenstad, L. M., Seewald, R. C., Cornelisse, L. E., and Shantz, J. (**1999**). "Comparison of linear gain and wide dynamic range compression hearing aid circuits: aided speech perception measures," Ear Hear. **20**, 117–26.

Kacelnik, O., Nodal, F. R., Parsons, C. H., and King, A. J. (**2006**). "Training-induced plasticity of auditory localization in adult mammals," PLoS biology **4**, 627–638.

Kaernbach, C. (**1991**). "Simple adaptive testing with the weighted up-down method," Percept. Psychophys. **49**, 227–9.

Kalluri, S., and Edwards, B. (**2007**). "Impact of hearing impairment and hearing aids on benefits due to binaural hearing," 19th International Congress on Acoustics (Madrid, Spain).

Kalluri, S., Eiler, C., and Edwards, B. (**2007**). "The benefit of binaural hearing to spatial perception in normal and hearing impaired listeners," California Academy of Audiology (Long Beach, CA), Vol. 45.

Keidser, G., Rohrseitz, K., Dillon, H., Hannacher, V., Carter, L., Rass, U., and Convery, E. (**2006**). "The effect of multi-channel wide dynamic range compression, noise reduction, and the directional microphone on horizontal localization performance in hearing aid wearers," Int. J. Audiol. **45**, 563–579.

Kidd, G., Arbogast, T. L., Mason, C. R., and Gallun, F. J. (**2005**). "The advantage of knowing where to listen," J. Acoust. Soc. Am. **118**, 3804–3815.

Kidd, G., Mason, C. R., and Gallun, F. J. (**2005**). "Combining energetic and informational masking for speech identification," J. Acoust. Soc. Am. **118**, 982.

Kochkin, S. (**2010**). "MarkeTrak VIII : Consumer satisfaction with hearing aids is slowly increasing," The Hearing Journal **63**, 19–27.

Kochkin, S. (**2011**). "MarkeTrak VIII: Patients report improved quality of life with hearing aid usage," The Hearing Journal **64**, 25–32.

Koehnke, J., Culotta, C. P., Hawley, M. L., and Colburn, H. S. (**1995**). "Effects of reference interaural time and intensity differences on binaural performance in listeners with normal and impaired hearing," Ear Hear. **16**, 331–53.

Koehnke, J., and Durlach, N. I. (**1989**). "Range effects in the identification of lateral position," J. Acoust. Soc. Am. **86**, 1176–8.

Levitt, H. (**1971**). "Transformed Up-Down Methods in Psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.

Macmillan, N. A., Rotello, C. M., and Miller, J. O. (**2004**). "The sampling distributions of Gaussian ROC statistics," Perception & psychophysics **66**, 406–21.

Markides, A. (**1982**). "The effectiveness of binaural hearing aids," Scandinavian audiology. Supplementum **15**, 181–96.

Marrone, N., Mason, C. R., and Kidd, G. (**2008**). "Evaluating the benefit of hearing aids in solving the cocktail party problem," Trends Amplif. **12**, 300–15.

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

Marrone, N., Mason, C. R., and Kidd, G. (**2008**). "The effects of hearing loss and age on the benefit of spatial separation between multiple talkers in reverberant rooms," J. Acoust. Soc. Am. **124**, 3064–3075.

Mencher, G. T., and Davis, A. (**2006**). "Bilateral or unilateral amplification: Is there a difference? A brief tutorial," Int. J. Audiol. **45**, S3–S11.

Mills, A. W. (**1958**). "On the Minimum Audible Angle," J. Acoust. Soc. Am. **30**, 237–246.

Monaghan, J. J. M., Krumbholz, K., and Seeber, B. U. (**2013**). "Factors affecting the use of envelope interaural time differences in reverberation," J. Acoust. Soc. Am. **133**, 2288–300.

Moore, B. B. C. J. (**2008**). "The choice of compression speed in hearing aids: Theoretical and practical considerations and the role of individual differences," Trends Amplif. **12**, 103–12.

Moore, B. C. (**1996**). "Perceptual consequences of cochlear hearing loss and their implications for the design of hearing aids," Ear Hear. **17**, 133–61.

Moore, B. C. J. (**2007**). *Cochlear Hearing Loss: Physiological, Psychological and Technical Issues* (Wiley, Chichester, UK), pp. 1–344.

Moore, B. C. J., Füllgrabe, C., and Stone, M. A. (**2010**). "Effect of spatial separation, extended bandwidth, and compression speed on intelligibility in a competing-speech task," J. Acoust. Soc. Am. **128**, 360–71.

Moore, B. C. J., and Glasberg, B. R. (**2004**). "A revised model of loudness perception applied to cochlear hearing loss," Hear. Res. **188**, 70–88.

Moore, B. C., Johnson, J. S., Clark, T. M., and Pluvinage, V. (**1992**). "Evaluation of a dual-channel full dynamic range compression system for people with sensorineural hearing loss," Ear Hear. **13**, 349–370.

Musa-Shufani, S., Walger, M., von Wedel, H., and Meister, H. (**2006**). "Influence of dynamic compression on directional hearing in the horizontal plane," Ear Hear. **27**, 279–285.

Nabelek, A. K., and Pickett, J. M. (**1974**). "Monaural and binaural speech perception through hearing aids under noise and reverberation with normal and hearing-impaired listeners," Journal of speech and hearing research **17**, 724–39.

Neher, T., Behrens, T., Carlile, S., Jin, C., Kragelund, L., Petersen, A. S., and Schaik, A. van (**2009**). "Benefit from spatial separation of multiple talkers in bilateral hearing-aid users: Effects of hearing loss, age, and cognition," Int. J. Audiol. **48**, 758–74.

Neher, T., Laugesen, S., Søgaard Jensen, N., and Kragelund, L. (**2011**). "Can basic auditory and cognitive measures predict hearing-impaired listeners' localization and spatial speech recognition abilities?," J. Acoust. Soc. Am. **130**, 1542.

Noble, W., Byrne, D., and Ter-Horst, K. (**1997**). "Auditory localization, detection of spatial separateness, and speech hearing in noise by hearing impaired listeners," J. Acoust. Soc. Am. **102**, 2343–52.

Noble, W., and Gatehouse, S. (**2006**). "Effects of bilateral versus unilateral hearing aid fitting on abilities measured by the Speech, Spatial, and Qualities of Hearing Scale (SSQ)," Int. J. Audiol. **45**, 172–81.

Noble, W., Ter-Horst, K., and Byrne, D. (**1995**). "Disabilities and handicaps associated with impaired auditory localization," J. Am. Acad. Audiol. **6**, 129–40.

Patterson, R. D., Allerhand, M. H., and Giguère, C. (**1995**). "Time-domain modeling of peripheral auditory processing: a modular architecture and a software platform," J. Acoust. Soc. Am. **98**, 1890–4.

Rakerd, B., and 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**, 3052–63.

Robillard, T., and Gillain, M. (**1996**). "Hearing aids: statistical study and satisfaction survey of patients in an ENT practice," Acta Otorhinolaryngol. Belg. **50**, 115–20.

Ruggles, D., and Shinn-Cunningham, B. (**2011**). "Spatial Selective Auditory Attention in the Presence of Reverberant Energy: Individual Differences in Normal-Hearing Listeners," J. Assoc. Res. Otolaryngol. **12**, 395–405.

Schwartz, A. H., and Shinn-Cunningham, B. G. (**2013**). "Effects of dynamic range compression on spatial selective auditory attention in normal-hearing listeners," J. Acoust. Soc. Am. **133**, 2329–2339.

Schwartz, A., McDermott, J. H., and Shinn-Cunningham, B. (**2012**). "Spatial cues alone produce inaccurate sound segregation: The effect of interaural time differences," J. Acoust. Soc. Am. **132**, 357–68.

Searle, C. L., Braida, L. D., Davis, M. F., and Colburn, H. S. (**1976**). "Model for auditory localization," J. Acoust. Soc. Am. **60**, 1164–75.

Sherbecoe, R. L., and Studebaker, G. A. (**2004**). "Supplementary formulas and tables for calculating and interconverting speech recognition scores in transformed arcsine units," Int. J. Audiol. **43**, 442–8.

Shinn-Cunningham, B. (**2000**). "Adapting to remapped auditory localization cues: A decision-theory model," Percept. Psychophys. **62**, 33–46.

Shinn-Cunningham, B. G. (**2005**). "Influences of spatial cues on grouping and understanding sound," Proceedings of the Forum Acusticum (Budapest, Hungary).

Shinn-Cunningham, B. G. (**2008**). "Object-based auditory and visual attention," Trends Cog. Sci. **12**, 182–186.

Shinn-Cunningham, B. G., and Best, V. (**2008**). "Selective attention in normal and impaired hearing," Trends Amplif. **12**, 283–99.

Shinn-Cunningham, B. G., Durlach, N. I., and Held, R. M. (**1998**). "Adapting to supernormal auditory localization cues I Bias and resolution," J. Acoust. Soc. Am. **103**, 3656–66.

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

Shinn-Cunningham, B. G., Schickler, J., Kopco, N., and Litovsky, R. (**2001**). "Spatial unmasking of nearby speech sources in a simulated anechoic environment," J. Acoust. Soc. Am. **110**, 1118–1129.

Simon, H. J., and Aleksandrovsky, I. (**1997**). "Perceived lateral position of narrow-band noise in hearing-impaired and normal-hearing listeners under conditions of equal sensation level and sound-pressure level," J. Acoust. Soc. Am. **102**, 1821–6.

Slaney, M. (**1998**). "Auditory Toolbox,". Retrieved June 13, 2013, from https://engineering.purdue.edu/~malcolm/interval/1998-010/

Stecker, G. C., and Hafter, E. R. (**2002**). "Temporal weighting in sound localization," J. Acoust. Soc. Am. **112**, 1046–1057.

Stone, M. A., and Moore, B. C. (**1992**). "Syllabic compression: effective compression ratios for signals modulated at different rates," Brit. J. Audiol. **26**, 351–61.

Stone, M. A., and Moore, B. C. J. (**2003**). "Effect of the speed of a single-channel dynamic range compressor on intelligibility in a competing speech task," J. Acoust. Soc. Am. **114**, 1023.

Stone, M. A., and Moore, B. C. J. (**2004**). "Side effects of fast-acting dynamic range compression that affect intelligibility in a competing speech task," J. Acoust. Soc. Am. **116**, 2311.

Stone, M. A., and Moore, B. C. J. (**2007**). "Quantifying the effects of fast-acting compression on the envelope of speech," J. Acoust. Soc. Am. **121**, 1654.

Stone, M. A., Moore, B. C. J., Fuellgrabe, C., and Hinton, A. C. (**2009**). "Multichannel Fast-Acting Dynamic Range Compression Hinders Performance by Young, Normal-Hearing Listeners in a Two-Talker Separation Task," J. Audio. Eng. Soc. **57**, 532–546.

Treisman, A., and Gelade, G. (**1980**). "A Feature-Integration Theory of Attention," Cognitive Psychology **12**, 97–136.

Verschuure, J., Maas, A. J., Stikvoort, E., de Jong, R. M., Goedegebure, A., and Dreschler, W. A. (**1996**). "Compression and its effect on the speech signal," Ear Hear. **17**, 162–75.

Whitmer, W. M., Seeber, B. U., and Akeroyd, M. A. (**2012**). "Apparent auditory source width insensitivity in older hearing-impaired individuals," J. Acoust. Soc. Am. **132**, 369–79.

Whitmer, W. M., Seeber, B. U., and Akeroyd, M. A. (**2013**). "Measuring the apparent width of auditory sources in normal and impaired hearing," Advances in experimental medicine and biology **787**, 303–10.

Wiggins, I. M., and Seeber, B. U. (**2011**). "Dynamic-range compression affects the lateral position of sounds," J. Acoust. Soc. Am. **130**, 3939–3953.

Wiggins, I. M., and Seeber, B. U. (**2012**). "Effects of dynamic-range compression on the spatial attributes of sounds in normal-hearing listeners," Ear Hear. **33**, 399–410.

Wiggins, I. M., and Seeber, B. U. (**2013**). "Linking dynamic-range compression across the ears can improve speech intelligibility in spatially separated noise," J. Acoust. Soc. Am. **133**, 1004–16.

Woods, D. L., Alain, C., and Ogawa, K. H. (**1998**). "Conjoining auditory and visual features during high-rate serial presentation: processing and conjoining two features can be faster than processing one," Perception & psychophysics **60**, 239–49.

Zurek, P. (**1993**). "Binaural advantages and directional effects in speech intelligibility," In G. Studebaker and I. Hochberg (Eds.), Acoustical Factors Affecting Hearing Aid Performance (Allyn and Bacon, Boston, MA), pp. 255–276.