2022

Modulation informational masking

https://hdl.handle.net/2144/44408

BOSTON UNIVERSITY

SARGENT COLLEGE OF HEALTH AND REHABILITATION SCIENCES

Dissertation

MODULATION INFORMATIONAL MASKING

by

CHRISTOPHER CONROY

B.M., Berklee College of Music, 2011

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
2022

Approved by

First Reader
Gerald Kidd, Jr., Ph.D.
Professor of Speech, Language & Hearing Sciences

Second Reader
Virginia Best, Ph.D.
Research Associate Professor of Speech, Language & Hearing Sciences

Third Reader
Emily Buss, Ph.D.
Professor of Otolaryngology/Head and Neck Surgery
University of North Carolina – School of Medicine

Fourth Reader
Richard L. Freyman, Ph.D.
Professor of Communication Disorders
University of Massachusetts – Amherst

ACKNOWLEDGMENTS

The work reported in this dissertation was supported by NIH-NIDCD grants No. R01DC004545 and F31DC019819. It could not have been completed without the support and guidance of Gerald Kidd, Jr. Emily Buss made substantial contributions to all chapters; Virginia Best and Richard L. Freyman provided valuable comments at various stages of their development. The members of the Psychoacoustics Laboratory, and, in particular, Christine R. Mason, also helped bring this work to fruition.

MODULATION INFORMATIONAL MASKING

CHRISTOPHER CONROY

Boston University Sargent College of Health and Rehabilitation Sciences, 2022

Major Professor: Gerald Kidd, Jr., Ph.D., Professor of Speech, Language & Hearing Sciences

ABSTRACT

The relatively slow variations in amplitude across time (amplitude modulation or AM) inherent to many behaviorally relevant sounds are of fundamental importance to auditory perception and communication. This is true for listeners with normal hearing, listeners with hearing loss, and, in particular, for listeners with cochlear implants, for whom limitations on frequency resolution imposed by current-generation implants make AM cues the dominant source of auditory information in some settings. As such, a core problem in contemporary hearing research is determining how AM cues are processed, in particular in multisource listening environments in which multiple AM components (peaks in the AM spectrum) associated with multiple, competing sources of sound occur simultaneously, and therefore must be processed selectively. While past research has revealed many of the low-level sensory mechanisms that mediate such processing, relatively little is known about the high-level, nonsensory factors involved. This dissertation reports the results of three studies that were designed to yield a better understanding of one such nonsensory factor; namely, listener uncertainty. The first study demonstrated that listener uncertainty regarding the AM spectrum of a masker (AM-rate uncertainty) can have an adverse effect on the detectability of target AM. The other two studies identified stimulus and psychological factors that can both produce and reduce this effect. In all three studies, psychophysical techniques that long have been used to investigate the effects of uncertainty in the context of auditory informational masking (IM) were adapted and applied to the study of AM-rate uncertainty for the first time, yielding insights into what we will call IM in the AM domain, or “modulation IM.” Taken together, the results shed new light on how uncertainty in general, and AM-rate uncertainty in particular, affects auditory perception and communication in the types of dynamic, multisource listening environments that characterize everyday life.

TABLE OF CONTENTS

LIST OF FIGURES

1. GENERAL INTRODUCTION

The basic problem that motivates the work described in this dissertation is broad, and it is one that is faced by listeners of all stripes on an almost daily basis: it is the problem of selectively attending to, and extracting information from, a sound of interest (target) along a particular acoustic dimension when that sound is embedded in an acoustic background (masker) that varies dynamically and unpredictably from one moment to the next. The crux of this problem is twofold. First, the dynamic and unpredictable changes in the acoustics of the masker may induce uncertainty in the listener, and this uncertainty may, in turn, make selective attention to, and information extraction from, the target much more difficult than it otherwise might be (cf. Tanner, 1961). Second, these changes may occur along any one of a variety of acoustic dimensions, or, more likely, along multiple dimensions simultaneously, and thus uncertainty may manifest with respect to any one or all of these dimensions (again, cf. Tanner, 1961), greatly compounding the selective attention and information extraction difficulties experienced by the listener.

For almost 50 years (Pollack, 1975), the dominant framework within auditory psychophysics for probing the effects of uncertainty in this context (i.e., multisource listening) has been that of informational masking (IM). One view of IM is that it is a type of masking that reflects interference in the selective processing of a target along a particular acoustic dimension that results from uncertainty induced by dynamic or unpredictable changes in a masker along that same acoustic dimension. Under this view, IM often is conceptualized as reflecting a failure of selective attention of some sort, consequent to the uncertainty induced by the masker (cf. Durlach et al., 2003a; Watson, 2005; Kidd et al., 2008). Thus, when IM occurs in a multisource mixture, it may reflect uncertainty-related selective-attention and information-extraction difficulties such as those noted above. Yet while much has been learned about the effects of uncertainty in such contexts from past research on IM, the multidimensional nature of the construct largely has been ignored, with the vast majority of past research on IM focused on how uncertainty arising from random variations in the frequency content of a masker affects selective attention to, and information extraction from, a target along that acoustic dimension; in other words, what we will call IM in the frequency domain, or "frequency IM" (cf. Durlach et al., 2003a). Considering the multidimensional nature of uncertainty, however, it is possible, and perhaps even likely, that the insights into uncertainty that have been obtained from past research on frequency IM may apply along other dimensions as well—dimensions along which uncertainty regarding random variations in a masker similarly may manifest. To more fully understand how uncertainty bears on auditory perception and communication in the types of listening environments that characterize everyday life, therefore, it is necessary to apply these insights more broadly. In this dissertation, we describe work that was designed to do just that and focus on how uncertainty arising from random variations in the amplitude-modulation (AM) content of a masker affects selective attention to, and information extraction from, a target along that acoustic dimension; in other words, what we will call IM in the modulation domain, or "modulation IM."

To help put the relevance of this research in context, consider, for a moment, the fundamental importance of AM cues to auditory perception and communication in everyday life, and thus the potentially devastating effects that an inability to access these cues due to modulation IM could have on real-world hearing. Nearly every sound that we encounter in our daily lives, including speech, music, and most environmental sounds, exhibit relatively slow variations in amplitude across time, or AM. In the case of speech, AM cues can be the primary carriers of intelligible speech information (e.g., Shannon et al., 1995). In the case of nonspeech sounds, they can provide crucial clues to sound-source identity and meaning (e.g., McDermott and Simoncelli, 2011). In everyday life, sounds of interest almost never occur in acoustic isolation; rather, they are embedded in an acoustic background of some kind. Given the prevalence of AM in the environment, this means that the particular (target) AM cues that subserve speech intelligibility, sound-source identification, and myriad other important auditory functions (e.g., source separation; e.g., Grimault et al., 2002) are almost always embedded in a background that exhibits (masker) AM cues of its own. Furthermore, in such situations, the AM cues associated with the target and those associated with the masker are often present in the same (or similar) frequency region(s), resulting in a complex amplitude envelope comprising multiple AM components (peaks in the AM spectrum). Consequently, the ability to selectively attend to, and extract information from, particular regions of the AM spectrum and ignore others likely is fundamental to auditory perception and communication in the real world. Considering again the fact that the types of listening environments that characterize everyday life not only contain multiple sources of sound but are also dynamic, meaning that the AM content both of the target and the masker may be constantly changing—often in unpredictable ways—the potential significance of modulation IM to real world hearing presents itself.

Yet, as it stands, modulation IM is a more or less novel area of inquiry. Indeed, save for a few of notable exceptions (e.g., Wright and McFadden, 1990; Richards et al., 1998), the effects of AM-rate uncertainty *per se* have received relatively little attention in the psychophysical literature on AM processing and perception. As such, the experiments reported in this dissertation were designed to characterize the basic aspects of modulation IM (i.e., the various factors that produce and reduce it) to get a handle on the basic phenomenon. This meant using highly simplified, nonspeech stimuli (i.e., modulated tones and noises, "complex" amplitude envelopes comprising relatively few components, etc.) and simple, psychophysical detection tasks. In this sense, the research reported in this dissertation is similar to the foundational research on frequency IM (see, e.g., the work of Watson and colleagues, Neff and colleagues, Lutfi and colleagues, and Kidd and colleagues; see citations in Durlach et al., 2003a; Watson, 2005; Kidd et al., 2008), which used simple, nonspeech stimuli (e.g., tones and multitone maskers) and simple, psychophysical detection and discrimination tasks as tools to elucidate the range of stimulus and psychological factors that can both produce and reduce frequency IM. The potential significance of the results to more complex stimuli and listening situations, however, is discussed briefly in Chapter 5.

The remainder of this dissertation is structured as follows[1]: In Chapter 2, an initial

---

[1] Please note that the three "content" chapters that comprise the bulk of this dissertation (i.e., Chapters 2–4) were written as stand-alone journal articles. As such, there is considerable overlap between them in terms of the material that is covered, in particular, in their Introduction and Discussion sections.

"proof-of-concept" study of modulation IM is reported, which provides a more complete development of the theoretical underpinnings of modulation IM, presents empirical data in support of its existence, and offers various potential interpretations of the effect. The key finding reported in Chapter 2 is that uncertainty induced by random variations in the AM content of a masker can have an adverse effect on the detectability of target AM. This, we will argue, is modulation IM. In Chapters 3 and 4, we take this basic finding as our point of departure and examine the extent to which different stimulus and psychological factors that are known to contribute to frequency IM contribute to modulation IM as well. In Chapter 3, we examine whether contextual cues to certain features of an otherwise highly uncertain stimulus AM spectrum can be used by listeners to obtain a release from modulation IM by deploying central, attentional mechanisms. In Chapter 4, by contrast, we examine whether stimulus manipulations intended to yield a relatively "automatic" perceptual segregation of the target AM can provide a release from modulation IM in a mixture. In Chapter 5, we provide a summary of the results of Chapters 2-4, as well as a general discussion of their implications. Taken together, the results highlight important similarities, but also important differences, between different types of IM arising from listener uncertainty induced by dynamic or unpredictable changes in a masker along different acoustic dimensions. As such, they shed new light on how uncertainty in general, and AM-rate uncertainty in particular, affect auditory perception and communication in everyday life.

Little effort was made to reduce such overlap because our primary concern in writing each chapter was its completeness as a journal article when excised from this dissertation. Similarly, important terms and abbreviations are reintroduced in each chapter, where necessary.

2. INFORMATIONAL MASKING IN THE MODULATION DOMAIN[2]

2.1. Abstract

Uncertainty regarding the frequency spectrum of a masker can have an adverse effect on the ability to focus selective attention on a target frequency channel, yielding informational masking (IM). This study sought to determine if uncertainty regarding the modulation spectrum of a masker can have an analogous adverse effect on the ability to focus selective attention on a target modulation channel, yielding IM in the modulation domain, or “modulation IM.” A single-interval, two-alternative forced-choice (yes-no) procedure was used. The task was to detect 32-Hz target sinusoidal amplitude modulation (SAM) imposed on a broadband-noise carrier in the presence of masker SAM imposed on the same carrier. Six maskers, spanning the range 8 to 128 Hz in half-octave steps, were tested, excluding those that fell within a two-octave protected zone surrounding the target. Psychometric functions (*d’* versus target-modulation depth) were measured for each masker under two conditions: a fixed (low-uncertainty/low-IM) condition, in which the masker was the same on all trials within a block, and a random (high-uncertainty/high-IM) condition, in which it varied randomly from presentation-to-presentation. Thresholds and slopes extracted from the psychometric functions differed markedly between the conditions. These results are consistent with the idea that IM occurs in the modulation domain.

---

[2] This chapter is a slightly edited version of an article that was previously published in the *Journal of the Acoustical Society of America* (Conroy and Kidd, 2021).

2.2. Introduction

There is considerable psychophysical evidence to support the conceptual model of amplitude modulation processing as mediated by an array of channels tuned to modulation frequency (e.g., Houtgast, 1989; Bacon and Grantham, 1989; Dau et al., 1997; Dau et al., 1999; Ewert and Dau, 2000; Ewert et al., 2002; Ewert and Dau, 2004; Sęk and Moore, 2003; Wojtczak and Viemeister, 2005; Moore et al., 2009; Wojtczak et al., 2011; Wojtczak, 2011; Sęk et al., 2015; Füllgrabe et al., 2021). The function of this putative array of channels is to decompose a stimulus comprising complex modulations into its constituent modulation frequencies, thereby enabling internal access to different regions of its amplitude-modulation spectrum.

Psychophysical modulation masking experiments illustrate this notion. In a typical modulation masking experiment, the listener's task is to detect the presence of target modulation—for example, sinusoidal amplitude modulation (SAM)—imposed on a carrier in the presence of masker modulation imposed on the same carrier. A modulation masking pattern is obtained by measuring target-SAM detection thresholds as a function of the modulation frequency of the masker SAM, with the modulation depth of the masker held fixed (e.g., Strickland and Viemeister, 1996; Ewert et al., 2002; Wojtczak, 2011; Sęk et al., 2015). What is typically found is that the modulation depth of the target at detection threshold is highest (poorest) for maskers that are close to the target in terms of modulation frequency, and lowest for those that are remote, with a gradual transition in between. In other words, a peaked masking pattern emerges, with a peak centered at or near the modulation frequency of the target.

Assuming detection via modulation channels, the interpretation of this pattern of results is straightforward: the peak in the masking pattern (i.e., modulation masking) reflects competition between the target SAM and masker SAM within a single, attended-to modulation channel (cf. Dau et al., 1997; Dau et al., 1999; Ewert and Dau, 2000; Ewert et al., 2002; Ewert and Dau, 2004; Wojtczak, 2011). When the target and masker are close in terms of modulation frequency, large amounts of masking occur because the masker strongly stimulates the attended-to channel, obscuring weak targets and raising detection thresholds in turn. When the target and the masker are remote in terms of modulation frequency, minimal masking occurs because the masker only weakly stimulates the attended-to channel—strongly driving some flanking channel instead—and thus can easily be ignored by the listener when making a detection decision. In the frequency domain, masking that results from competition between a target and a masker within a single frequency channel is often referred to as energetic masking (EM; cf. Durlach et al., 2003a). Thus, by way of analogy, modulation masking such as that described above can be interpreted as EM in the modulation domain inasmuch as it reflects competition between the target SAM and masker SAM within a single, attended-to modulation channel (cf. Durlach et al., 2003a; Sheft and Yost, 2007).

An interpretation of modulation masking in terms of EM has intuitive appeal, and accords well with the results of many modulation masking experiments (see references above). Yet, in the frequency domain, it has long been known that there are certain masked detection conditions under which EM alone is insufficient to explain performance. For example, when masker-frequency uncertainty is created in a masked pure-tone detection experiment by randomly varying the frequency spectrum of the masker from presentation-to-presentation (e.g., Neff and Green, 1987), large amounts of masking can occur that cannot be attributed solely to EM (for a review of many such experiments, see Kidd et al., 2008). Instead, the preponderance of masking in such cases likely arises because uncertainty regarding the frequency spectrum of the masker has an adverse effect on the ability to focus selective attention on the target frequency channel and ignore the uninformative flanking channel(s) on each trial (e.g., Lutfi, 1993; Neff et al., 1993; Allen and Wightman, 1995; Oh and Lutfi, 1998; Wright and Saberi, 1999; Richards et al., 2002; Tang and Richards, 2003; Lutfi et al., 2003b; Durlach et al., 2005). In this chapter, we report the results of an experiment designed to answer the following question: Does uncertainty regarding the modulation spectrum of a masker in a modulation masking experiment have an analogous adverse effect on the ability to focus selective attention on a target modulation channel?

The term informational masking (IM) is often used to refer to interference in detection performance that cannot be attributed to EM (cf. Durlach et al., 2003a). For a particular masking effect to qualify as IM, however, it is often stipulated that it must be the result of uncertainty arising from random or unexpected variations in a masker along one or more acoustic dimensions (e.g., Watson and Kelly, 1981; Neff and Green, 1987; Watson, 1987; Neff and Callaghan, 1988; Lutfi, 1990; Oh and Lutfi, 1998; Lutfi et al., 2003b), although other factors (e.g., target-masker similarity; Kidd et al., 1994; Neff, 1995; Kidd et al., 2002; Durlach et al., 2003b; Lee and Richards, 2011) may contribute as well. Thus, when masker-frequency uncertainty is created in a masked pure-tone detection experiment by randomly varying the frequency spectrum of the masker from presentation-to-presentation, the masking that results is considered to be due to IM. In this dissertation, we refer to it as IM in the frequency domain, or “frequency IM,” to denote the following: (1) the dimension along which the detection-relevant and irrelevant flanking channels are tuned is frequency; (2) the uncertainty that produces the masking is the result of random or unexpected variations in the frequency spectrum of the masker; and (3) the hypothesized consequence of the uncertainty is an inability to focus selective attention on the target frequency channel. Thus, swapping “frequency” for “modulation” in each of these criteria and considering modulation masking as EM in the modulation domain, the question posed in the preceding paragraph can be reformulated as follows: Is there such a thing as IM in the modulation domain, or “modulation IM”?

We are not unique in posing this question. Durlach et al. (2003a) noted that while the vast majority of past research on IM has focused on the frequency domain, it could easily occur in other, “nonfrequency domains” (pp. 2985) as well, insofar as channel-based processing is a factor therein. Indeed, they suggested the modulation domain as a possibility in a parenthetical remark (see pp. 2985), but never reported the results of such an experiment, or elaborated further. Sheft and Yost (2007) conceptualized modulation-detection-interference (MDI)—a phenomenon in which masker modulations in one carrier-frequency region interfere with the detection of target modulations in another—as an IM-like effect. The basis for their conceptualization was that the pattern of MDI that they observed in their study could not be explained in terms of competition between the target and masker modulations within modulation-tuned channels; that is, by EM in the modulation domain. Although their findings and interpretation do satisfy many of the criteria typically applied to IM and thus could be considered as modulation IM, they differ from the view proposed here in two important respects. First, our conceptualization of modulation IM depends on the central role of uncertainty and considers masker-modulation-frequency uncertainty resulting from random or unexpected variations in the modulation spectrum of a masker to be fundamental to modulation IM. Masker-modulation-frequency uncertainty was not an element of the design and interpretation of the experiments reported by Sheft and Yost. Second, we emphasize a theoretical account of IM in which the hypothesized consequence of uncertainty is an inability to focus selective attention on a target channel tuned in the domain in which the uncertainty exists. Although the empirical results reported by Sheft and Yost are consistent with such a failure, this was not the mechanism they invoked to explain their findings.

In the experiment reported here, target-SAM detection performance was measured in a modulation masking experiment very much like the “typical” modulation masking experiment described above only modified to include, in the key condition, masker-modulation-frequency uncertainty. Masker modulation-frequency uncertainty was created in a manner analogous to the way in which masker-frequency uncertainty is often created in masked pure-tone detection experiments concerned with frequency IM; namely, by randomly varying the modulation frequency of the masker SAM from presentation-to-presentation. Any adverse effect of uncertainty in this context was taken as evidence of modulation IM. Our working hypothesis was that, if evident, modulation IM would reflect an adverse effect of uncertainty on the ability to focus selective attention on the target modulation channel.

2.3. Methods

2.3.1. Listeners

Eight listeners (four males; 20-31 years; mean=23 years) participated. All listeners had pure-tone air-conduction thresholds within the normal range at octave frequencies from 250 to 8000 Hz. One listener was the author (CC); another was a researcher in our lab. Recruitment and use of human subjects protocols were approved by the Boston University Charles River Campus Institutional Review Board.

2.3.2. Apparatus

Stimuli were generated using MATLAB (MathWorks, Inc., Natick, MA), routed through a 24-bit sound card (RME HDSP 9632, Haimhausen, Germany), and presented monaurally to the listeners' left ears via a pair of headphones (Sennheiser HD280 pro, Wedemark, Germany). Listeners performed the task individually while seated in a double-walled, sound-treated booth (Industrial Acoustics Company, North Aurora, IL) equipped with a computer monitor, a keyboard, and a mouse.

2.3.3. Stimuli and procedures

The details of the experimental design followed closely those described by Durlach et al. (2005) in their study of frequency IM. A single-interval, two-alternative forced-choice (yes-no) procedure was used. The task was to detect target SAM (the target) of a fixed and known modulation frequency imposed on a broadband noise carrier with a 0.5 *a priori* probability on each trial. Simultaneous masker SAM (the masker) was imposed on the same noise carrier with a 1.0 *a priori* probability on each trial. The carrier was 500 ms and was different on each trial. The masker and, when present, the target were applied to the entire 500-ms duration. On masker-alone trials, the equation that described the waveform was

$$[1+m_m\sin(2\pi f_m t)]n(t) \tag{1}$$

where m_m was the modulation depth of the masker, f_m was the modulation frequency of the masker in Hz, and $n(t)$ was the noise carrier. On target+masker trials, the equation that described the waveform was

$$[1+m_m\sin(2\pi f_m t)+m_t\sin(2\pi f_t t)]n(t) \tag{2}$$

where m_t was the modulation depth of the target and f_t was the modulation frequency the target in Hz.

The modulation frequency of the target was 32 Hz. The modulation depth of the masker was 0.5, or -6 dB in terms of 20log(m_m). Six maskers were tested, spanning the range 8 to 128 Hz in half-octave steps. They were 8, 11, 16, 64, 91, and 128 Hz. Maskers that fell within a two-octave protected zone surrounding the target were excluded in order to minimize the effects of modulation masking (i.e., EM in the modulation domain) so that any effects of modulation IM could be more readily identified.

On each trial, the waveform described by either Equation 1 or 2 was bandpass filtered between 80 and 8000 Hz using an 8th order Butterworth bandpass filter, ramped using 50-ms cosine-squared onset-offset ramps, and scaled to an overall level of 50 dB SPL before presentation to the listener during a 500-ms observation interval. The observation interval was preceded by a 500-ms warning light (displayed on the in-booth monitor) and followed by a response period of unlimited duration. The listener's response ("target present," yes-no) was registered via mouse click on a labeled button displayed on the in-booth monitor. Once registered, no opportunity for corrections was provided and correct-answer feedback was displayed immediately on the monitor for 500 ms before the next trial commenced or the block of trials terminated.

Target-SAM detection performance was measured in two stimulus conditions: a fixed condition, denoted *F*, in which the masker was the same on all trials within a block, and a random condition, denoted *R*, in which the masker was selected at random on each trial from the set of six. The use of the same masker across trials in the *F* condition was assumed to preclude masker-modulation-frequency uncertainty (or at least that masker-modulation-frequency uncertainty was at a minimum) and thus the *F* condition was considered the low-uncertainty/low-IM reference condition. The presentation-to-presentation randomization of the masker in the *R* condition was intended to result in (or to increase the amount of) masker-modulation-frequency uncertainty (relative to the *F* condition) and thus the *R* condition was considered the high-uncertainty/high-IM comparison condition. It is worth reiterating here that when we say "the masker" and refer to its randomization we are referring to the imposed masker modulator. The noise carrier, which was unique on each trial, had inherent fluctuations of its own that also could have produced modulation masking (e.g., Dau et al., 1999) and/or uncertainty (cf. Spiegel and Green, 1982). Target-SAM detection performance was evaluated for each masker in each stimulus condition as a function of target modulation depth and quantified via the standard detection-theory index of sensitivity *d'*. Three target modulation depths spaced 5 dB apart were tested for each masker, with the specific depths chosen to ensure that a wide range of performance was obtained for each masker. For a given masker, the same three depths were tested in both the *F* and *R* conditions.

Prior to beginning data collection, each listener completed six adaptive tracks of unmasked target-modulation detection to gain familiarity with the target. A two-interval, two-alternative forced-choice procedure was used and correct-answer feedback was provided on all trials. The same stimuli were used during the familiarization period as were used during the experimental session save for the fact that a masker was never present.

During the experimental session, trials were completed in blocks of 50 and *F* and *R* blocks were completed in alternation, always beginning with an *F* block. In the *F* condition, the target depth was the same on all trials within a block. The three target depths that were tested for each masker were rank-ordered (highest to lowest) and, following Durlach et al. (2005), were tested in decreasing order across blocks to minimize any uncertainty associated with variability in target depth that might have arisen had we adapted on target depth or used the method of constant stimuli. A block of trials was completed for each masker at each rank-ordered value before proceeding to the next rank-ordered value for any other masker. The order (different for each listener) in which the six maskers were tested at each rank-ordered value was randomized across blocks. Note that this randomization procedure applied to the *F* condition only; in the *R* condition, the masker was selected at random on each trial. Moreover, in the *R* condition, the target-modulation depth was fixed within a block according to its rank-ordered value (as opposed to its absolute value, as in the *F* condition) and the three rank-ordered values were tested in decreasing order across blocks.

The data collection procedure described above yielded a 2×2 stimulus-response matrix (i.e., number of hits, misses, false alarms, and correct rejections) based on 50 trials for each combination of target depth, masker, and stimulus condition. It was completed in a single experimental session lasting roughly 1.5 hours. After this first experimental session was completed, it was repeated (on a different day, preceded by two additional unmasked target-modulation detection blocks for the purposes of refamiliarization with the target) to arrive at the final data set for each listener, i.e., a 2×2 stimulus-response matrix based on 100 trials for each combination of target depth, masker, and stimulus condition.

2.3.4. Analysis conditions

Only two stimulus conditions were tested, *F* and *R*. Following Durlach et al. (2005), however, the final data set for each listener in each stimulus condition was subjected to two separate analyses, yielding *d'* values for two pairs of analysis conditions. For the two "pooled" analysis conditions, F_p and R_p, all 2×2 stimulus-response matrices associated with a particular target depth and stimulus condition were pooled across maskers (i.e., the numerical values in each cell were summed across maskers) and d'_p was calculated on the basis of these pooled matrices. In the second pair of analysis conditions, detection performance was evaluated at the individual masker level. For these two "sorted" analysis conditions, F_s and R_s, the 2×2 stimulus-response matrices associated with each target depth were sorted by masker and stimulus condition and *d's* was calculated on the basis of these sorted matrices.

Regardless of the analysis condition, each 2×2 stimulus-response matrix was converted to a matrix of response probabilities (i.e., probability of a hit, miss, false alarm, and correction rejection), extreme probabilities were corrected using the "1/(2N)" rule (Hautus, 1995), and *d'* was computed via the standard equation for yes-no tasks (see, e.g., Macmillan and Creelman, 2005, pp. 8); that is,

$$d' = z(P_H) - z(P_{FA}) \qquad (3)$$

where $z(P_H)$ was the *z*-score of the probability of a hit and $z(P_{FA})$ was the *z*-score of the probability of a false alarm.

2.4. Results

Figure 2.1 shows group-mean d'_p plotted as a function of target modulation depth in dB for each of the two pooled analysis conditions (cf. Section II.D), F_p (filled symbols) and R_p (open symbols). Error bars give ±1 standard error of the mean (SEM). Also shown are group-mean psychometric functions for the same two conditions, F_p (solid line) and R_p (dotted line). The group-mean psychometric function for each condition was obtained by averaging the parameters (slope and y-intercept) of individual psychometric functions fit separately to the d'_p data for each listener in each condition. The individual psychometric functions were computed as linear least-squares fits to the d'_p data on the coordinates of d'_p versus target-modulation depth in dB. Following Durlach et al. (2005), threshold-level performance was defined as the target-modulation depth in dB at which the fitted psychometric function crossed d'_p=2; following Sęk et al. (2015), if the fitted function failed to cross d'_p=2 by a target-modulation depth of 0 dB, the threshold was set to 0 dB to acknowledge the constraint of overmodulation (referred to as "the 0-dB maximum threshold rule" in what follows). Individual and group-mean thresholds (top panel) and psychometric function slopes (bottom panel) are plotted in Figure 2.2. Note that, in Figure 2.2, the two listeners from our lab, the author (C.C.) and another researcher, are Observers 1 and 2, respectively.

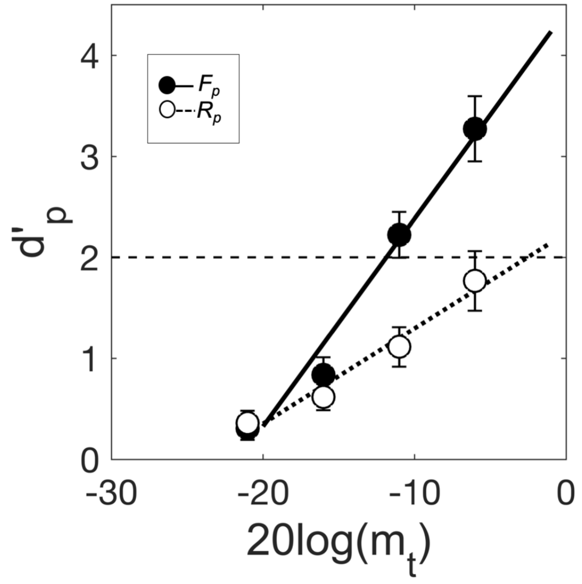

Figure 2.1. Group-mean detectability (d'_p) plotted as a function of target modulation depth in dB for each of the two pooled analysis conditions: the low-uncertainty/low-IM fixed, F_p (filled symbols), condition and the high-uncertainty/high-IM random, R_p (open symbols), condition. Error bars give ±1 standard error of the mean (SEM). Also shown are group-mean psychometric functions for the F_p (solid line) and R_p (dotted line) conditions (see text for details). The horizontal dashed line marks d'_p=2, or threshold-level performance.

Turning first to the thresholds: thresholds were higher (poorer) in the R_p condition than in the F_p condition for all listeners, indicating an adverse effect of masker-modulation-frequency uncertainty on target SAM detection. This, we argue, is modulation IM. Individual F_p thresholds ranged from -15 to -4 dB and the group-mean F_p threshold was -11 dB (SEM=1 dB). R_p thresholds ranged from -12 to 0 dB, with a group mean of -3 dB (SEM=2 dB). Because modulation masking (i.e., EM in the modulation domain) was equivalent in both the low-IM F_p and high-IM R_p conditions, the difference between each listener's F_p and R_p threshold ($R_p - F_p$) provides an index of modulation IM analogous to that which is typically used to index frequency IM (cf. Lutfi, 1990; Durlach et al., 2003a). In general, a positive threshold difference indicates the presence of modulation IM and larger threshold differences indicate greater susceptibility to modulation IM, whereas smaller threshold differences indicate greater resilience (with the caveat that the 0-dB maximum threshold rule put a cap on threshold differences in some cases; cf. Figure 2.2). Threshold differences ranged from 3 to 12 dB. The group-mean threshold difference was 8 dB (SEM=1 dB). A one-tailed t-test confirmed that the mean threshold difference was significantly greater than zero [$t(7) = 6.20$, $p < .001$], suggesting that modulation IM was a significant factor in the R_p condition.

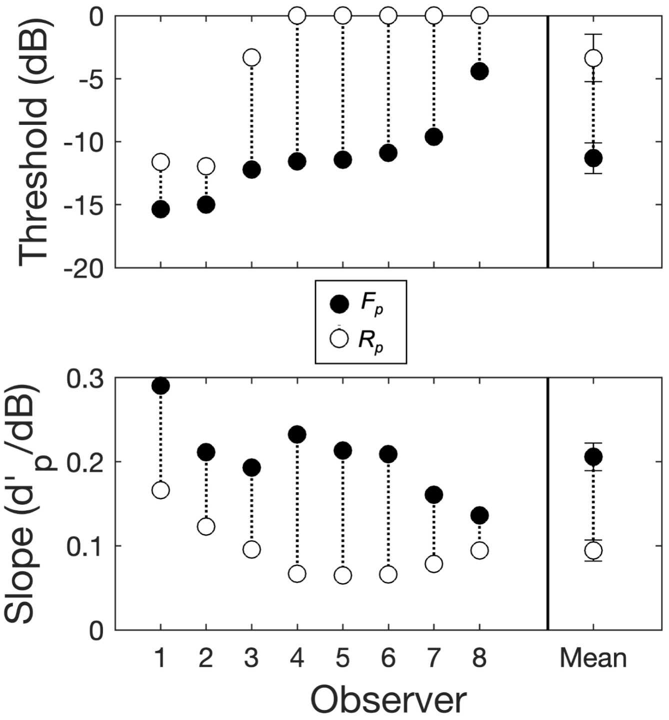

Figure 2.2. Individual and group-mean thresholds (top panel) and psychometric function slopes (bottom panel) in each of the two pooled analysis conditions, F_p and R_p. Error bars give ±1 SEM.

Turning now to the d'_p psychometric functions: consistent with the results of numerous past studies of frequency IM (e.g., Neff and Callaghan, 1988; Kidd et al., 1994; Wright and Saberi, 1999; Lutfi et al., 2003b; Tang and Richards, 2003; Kidd et al., 2003; Durlach et al., 2005), psychometric functions were shallower in the high-IM R_p condition than in the low-IM F_p condition, again for all listeners. The slopes of the F_p functions ranged from 0.14 to 0.29 d'_p/dB and the group-mean F_p slope was 0.21 d'_p/dB. The slopes of the R_p functions ranged from 0.06 to 0.17 d'_p/dB and the group-mean R_p slope was 0.10 d'_p/dB. Unlike the threshold-difference metric, the ratio of psychometric function slopes (F_p/R_p) for each listener provides an index of modulation IM (cf. Durlach et al., 2005) uncontaminated by the 0-dB maximum threshold rule. A slope ratio $\neq 1$

indicates an effect of modulation IM. Slope ratios ranged from 1.44 to 3.47 and the group-mean slope ratio was 2.36 (SEM=0.29). A two-tailed t-test confirmed that slope ratios were significantly greater than one [$t(7) = 4.76$, $p < .01$], suggesting that that modulation IM yielded a significant decrease in the slope of the psychometric function for target SAM detection.

The pooled analysis presented above was used to examine the effect of uncertainty at the masker ensemble level; the sorted analysis presented below, on the other hand, was used to examine the effect of uncertainty at the individual masker level (cf. Section II.D). As with the d'_p data, the d'_s data for each listener, each masker, and each condition were fit with a straight-line psychometric function on the coordinates of d'_s versus target-modulation depth in dB and thresholds and slopes were extracted from the fitted psychometric functions as described above; group-mean psychometric functions were obtained for each masker in each condition, again, by averaging the parameters (slope and y-intercept) of the individual listener fits. Figure 2.3 shows group-mean d'_s plotted as a function of target modulation depth in dB as well as the group-mean psychometric function for each masker (different panels, masker modulation frequency inset) in each of the two sorted analysis conditions, F_s (filled symbols, solid lines) and R_s (open symbols, dotted lines). Error bars give ±1 SEM.

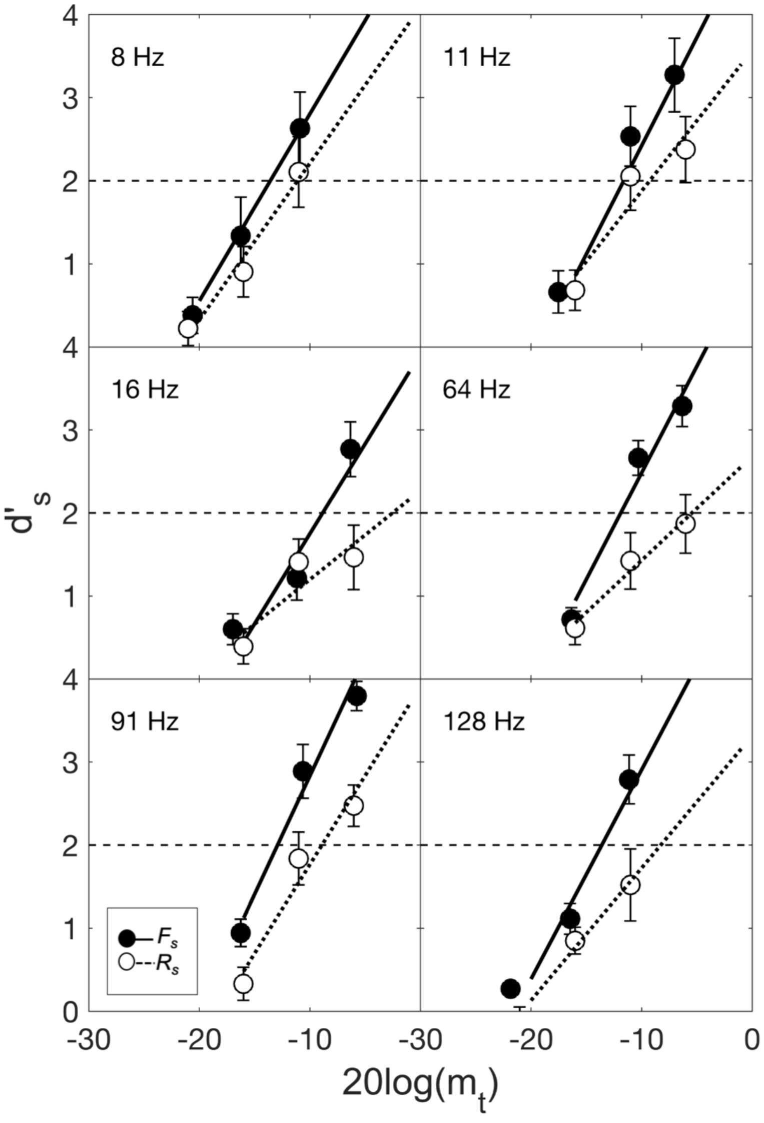

Figure 2.3. Group mean *d'*$_s$ plotted as a function of target modulation depth in dB for each masker (panels, masker modulation frequency inset) in each of the two sorted analysis conditions, F_s (filled symbols) and R_s (open symbols). Error bars show ±1 SEM. Note that the symbols have been jittered along the abscissa slightly to improve clarity. Also shown are group mean psychometric functions for each masker in the F_s (solid lines) and R_s (dotted lines) conditions. The horizontal dashed line in each panel marks d'_p=2, or threshold-level performance.

Consistent with the d'_p results above, the d'_s results in Figure 2.3 show that masker-modulation-frequency uncertainty had a generally adverse effect on target-SAM detection performance across maskers, yielding both a positive threshold difference

($R_s - F_s$) and a slope ratio (F_s/R_s) >1 in all cases. This is illustrated more clearly in Figure 2.4, which shows group-mean thresholds (left panel) and psychometric function slopes (right panel) for each masker plotted as a function of the modulation frequency of the masker in both the F_s (filled symbols, solid lines) and R_s (open symbols, dotted lines) conditions. Error bars give ±1 SEM. The vertical displacement of the F_s and R_s data in each panel of Figure 2.4 illustrates the effect of uncertainty on thresholds (left panel) and slopes (right panel): R_s symbols above F_s symbols in the left panel of Figure 2.4 indicate positive threshold differences, whereas R_s symbols below F_s symbols in the right panel of Figure 2.4 indicate slope ratios >1.

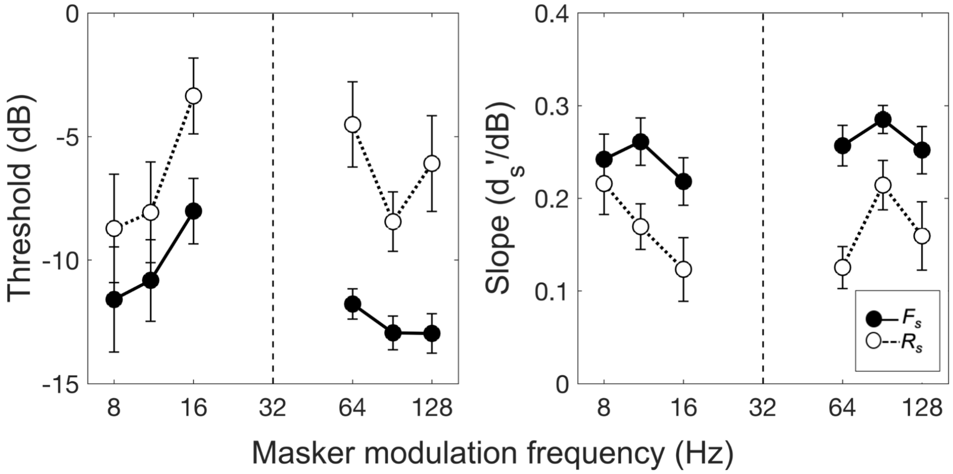

Figure 2.4. Group-mean thresholds (left panel) and psychometric function slopes (right panel) for each masker plotted as a function of the modulation frequency of the masker in each of the two sorted analysis conditions, F_s (filled symbols, solid lines) and R_s (open symbols, dotted lines). Error bars give ±1 SEM. The vertical dashed line marks 32 Hz, the modulation frequency of the target.

A two-way repeated-measures analysis of variance (ANOVA) performed on the thresholds revealed that both the main effects of uncertainty [F_s vs. R_s, $F(1,7) = 31.70$, p

< .001] and masker modulation frequency [six modulation frequencies, *F*(2.37,16.61) = 3.86, $p < .05$, Greenhouse-Geisser correction] were significant.[3] Notably, the interaction term (uncertainty×masker modulation frequency) approached, but did not reach, statistical significance at the $p < .05$ level [*F*(5,35) = 2.12, $p = .09$], indicating that threshold differences were comparable across masker modulation frequencies. The lack of a statistically significant interaction comes with a caveat, however, in that it may reflect, in part, the 0-dB maximum threshold rule, which limited the dynamic range available to measure threshold differences for some maskers, and, in particular, for the two maskers immediately flanking the target.[4] Indeed, visual inspection of the group-mean psychometric functions shown in Figure 2.3 suggests that threshold differences tended to be relatively large for the two maskers immediately flanking the target and that, without an "artificial" cap of 0 dB, threshold differences may have been larger still.

A second ANOVA performed on the psychometric function slopes indicated that, as expected, the main effect of uncertainty was significant [F_s vs. R_s, *F*(1,7) = 29.79, $p <$ .001]. Neither the main effect of masker modulation frequency [six modulation frequencies, *F*(3.09,21.61) = 2.08, $p > .05$, Greenhouse-Geisser correction] nor the

---

[3] A series of *post hoc* t-tests comparing thresholds among maskers (two-tailed, p < .05, Bonferroni correction) indicated that the main effect of masker modulation frequency was driven primarily by the elevated thresholds produced by the 16-Hz masker in both the F_s and R_s conditions. Specifically, the group-mean threshold for the 16-Hz masker averaged across the F_s and R_s conditions (-6 dB, SE=1 dB) was significantly higher than the group-mean threshold (again, averaged across conditions) for the 11-Hz masker (-9 dB, SE=1 dB), the 91-Hz masker (-11 dB, SE= 1 dB), and the 128-Hz masker (-10 dB, SE=1 dB). Moreover, the group-mean threshold for the 64-Hz masker (-8 dB, SE=1 dB) was significantly higher than that for the 91-Hz masker. No other comparisons reached statistical significance at the p<0.05 level.

[4] In the R_s condition, the 0-dB maximum threshold rule was required in 12 of 48 cases. Seven of those 12 cases were for either the 16-Hz or 64-Hz masker.

interaction between uncertainty and masker modulation frequency [$F(5,35) = 1.02$, $p > .05$] were significant, however, indicating that while masker-modulation-frequency uncertainty may have yielded flatter psychometric functions overall, the degree of flattening was similar across maskers.

2.4. Discussion

The purpose of this study was to answer a relatively simple question: Is there such a thing as IM in the modulation domain, or, what we have called, modulation IM? In short, the results suggest, Yes. Analyzed at the masker ensemble level (pooled analysis conditions), target-SAM detection thresholds were poorer, and psychometric functions were flatter, when masker-modulation-frequency uncertainty was high than when it was low (Figures 2.1 and 2.2). Analyzed at the individual masker level (sorted analysis conditions), masker-modulation-frequency uncertainty had a generally adverse effect on thresholds across maskers (Figures 2.3 and 2.4), yielding shallower psychometric functions as well (Figures 2.3 and 2.4). Because modulation masking (i.e., EM in the modulation domain) was equated across both the *F* and *R* conditions through the use of a protected zone and the same set of six maskers, we argue that these findings reflect the effects of modulation IM.

As noted in Section I, our working hypothesis was that, if evident, modulation IM would reflect an adverse effect of masker-modulation-frequency uncertainty on the ability to focus selective attention on a target modulation channel. The pattern of thresholds and slopes summarized above is generally consistent with this hypothesis. For example, if it is assumed that, in the *F* condition, listeners were able to focus their attention on the target modulation channel but that, in the *R* condition, were either distracted or confused by the masker on some trials and therefore attended to the "wrong" (i.e., nontarget) modulation channel instead, an increase in thresholds and decrease in psychometric function slopes in the high-IM *R* condition relative to the low-IM *F* condition would be expected (cf. Allen and Wightman, 1995; see also Hübner, 1993; Green, 1995). Insofar as all maskers were equally distracting/confusing, this account would predict a similar effect of uncertainty on both thresholds and slopes across maskers (i.e., comparable threshold differences and slope ratios across maskers) and therefore would be consistent both with the results on d'_p (Figures 2.1 and 2.2) and on d'_s (Figures 2.3 and 2.4). It also suggests an important role for target-masker confusions in producing modulation IM (cf. Sheft and Yest, 2007), very much like how target-masker confusions can be an important factor in producing frequency IM (e.g., Kidd et al., 1994; Neff, 1995; Kidd et al., 2002; Durlach et al., 2003b; Lee and Richards, 2011).

Of course, there are other interpretations, unrelated to attention, that could have yielded a similar pattern of thresholds and slopes across conditions. A sufficient increase, under uncertainty, in the variance of some putative "internal noise," for example, would be expected to produce a similar pattern, both in terms of the thresholds and slopes derived from the d'_p and d'_s data. Another interpretation, again, unrelated to attention, is that, in both the *F* and *R* conditions, listeners employed a detection strategy in which the presence of the target was determined via a "qualitative change in the character of the masker" (Neff and Green, 1987, pp. 412-413) resulting from the addition of the target.

The logic of this interpretation is as follows: in the *F* condition, the same masker sample was presented on all trials within a block. As such, the subjective "quality of the masker" (Neff and Green, 1987, pp. 410) was likely relatively stable across trials and thus listeners could have simply responded "target present" on those trials on which, due to the addition of the target, the subjective quality of the masker most strongly deviated from the expectation. Randomization of the masker in the *R* condition, by contrast, would have produced large variations in the subjective quality of the masker across trials—regardless of whether the target was added or not—rendering this strategy ineffective; large amounts of modulation IM, therefore, would be expected to occur.

To examine this possibility in more detail, a brief follow-up experiment was conducted. The stimuli (target, set of six maskers, carrier, etc.) and task were the same as in the main experiment but, instead of using a single-interval yes-no procedure, a two-interval, two-alternative forced-choice procedure was used. One each trial, one interval contained the target and one did not; the listener's task was to indicate which interval contained the target. Two listeners (two females; 21-22 years) who had not taken part in the main experiment participated. Three conditions were tested: (1) a fully fixed condition, in which the modulation frequency of the masker was the same on all trials and intervals within a block, (2) a between-trial randomization condition, in which the modulation frequency of the masker was selected at random on each trial but was fixed across intervals, and (3) a within-trial randomization condition, in which the modulation frequency of the masker was selected at random on each stimulus presentation. The idea was that a detection strategy based on a qualitative change in the character of the masker would be effective in both the fully fixed and between-trial randomization conditions due to the stability of the masker quality across trials and/or intervals in these conditions, but would be relatively ineffective in the within-trial randomization condition for the reason noted above. In all three conditions, the modulation depth of the masker was set at -6 dB (as in the main experiment) and the modulation depth of the target was set to the group-mean F_p threshold from the main experiment (-11 dB) in order to estimate *d'*.[5]

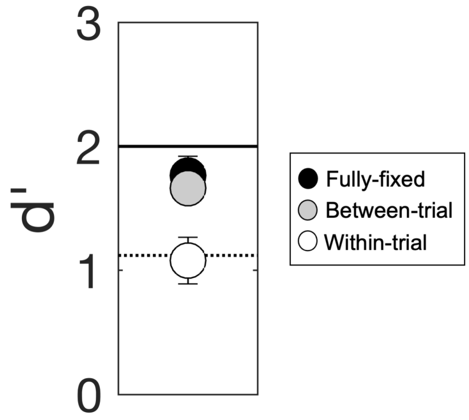

Figure 2.5. Results of the follow-up experiment. Mean *d'* for the two listeners in each of the three conditions: the fully fixed condition (filled black symbol), the between-trial randomization condition (filled grey symbol), and the within-trial randomization condition (open symbol). Error bars give ±1 SEM. The horizontal lines show d'_p values for the F_p (solid line) and R_p (dotted line) conditions of the main experiment extracted from the group-mean d'_p psychometric functions at the point at which the target-modulation depth (-11 dB) was the same as in the follow-up experiment.

Figure 2.5 shows the mean *d'* across the two listeners in each of the three conditions tested in the follow-up experiment: the fully fixed condition (filled black

---

[5] Proportion correct performance was measured for each of the three conditions over six blocks of 50 trials each. The proportion correct score for each block was converted to *d'* (see, e.g., Macmillan and Creelman, 2005, pp. 172) without respect to the particular masker sample that was used on each interval/trial (as in the pooled analysis conditions of the main experiment) and the final estimate of *d'* for each observer in each condition was taken as the mean of the *d'* estimates across the six blocks.

symbol), the between-trial randomization condition (filled grey symbol), and the within-trial randomization condition (open symbol). Error bars give ±1 SEM. For comparison, the horizontal lines show d'_p values for the F_p (solid line; d'_p=2) and R_p (dotted line; d'_p=1.12) conditions of the main experiment, extracted from the group-mean d'_p psychometric functions at the point at which the target-modulation depth (-11 dB) was the same as in the follow-up experiment (cf. Figure 2.1).

The first point to be made with respect to the results shown in Figure 2.5 is that, consistent with the results of the main experiment both in terms of trend and magnitude, masker-modulation-frequency uncertainty in the within-trial randomization condition (a condition analogous to the *R* condition of the main experiment) had an adverse effect on target modulation detection performance relative to the fully fixed condition (a condition analogous to the *F* condition of the main experiment): mean *d'* was 1.77 (SEM=0.15) in the fully fixed condition vs. 1.08 (SEM=0.19) in the within-trial randomization condition. Notably, however, there was effectively no difference in *d'* between the fully fixed and between-trial randomization conditions (the mean *d'* in the between-trial randomization condition was 1.67, SEM=0.03) indicating that fixing the modulation frequency of the masker across intervals of each two-interval forced-choice trial effectively eliminated modulation IM. This is consistent with the notion that detection of the target both here and in the main experiment could have been based on a change in subjective masker quality. The subjective quality judgements could have been based on a variety of cues (e.g., cues related to the temporal structure of the envelope; cf. Strickland and Viemeister, 1996; Viemeister et al., 2005) and thus this finding raises the more general possibility that modulation channels may not be required to explain our results.

The results of the follow-up experiment speak to another possible interpretation of the results of the main experiment; namely, that adaptation of the masker modulation (cf. Tansley and Suffield, 1983; Richards et al., 1997; Wojtczak and Viemeister, 2003) played a role. That is, it is possible that, in the *F* condition, long-term adaptation of the masker modulation resulting from the repeated presentation of the same masker across trials reduced the ability of individual masker samples to produce masking relative to when those same maskers occurred in a random-masker context (i.e., in the *R* condition). If that were the case, however, we would expect to see a difference in *d'* between the fully fixed and between-trial randomization conditions in the follow-up experiment (i.e., greater adaptation resulting from repeated presentations of the same masker across trials vs. only across intervals), whereas none was found (cf. Figure 2.5). It is possible however, that adaptation effects were effectively at their maximum following a single stimulus exposure, in which case no difference in *d'* between the fully fixed and between-trial randomization conditions would be expected. Indeed, whereas some past studies of modulation adaptation have used relatively long adaptors (i.e., on the order of minutes) in order to achieve adaptation effects (e.g., Tansley and Suffield, 1983; Richards et al., 1997; Wojtczak and Viemeister, 2003), results from modulation forward-masking experiments (e.g., Wojtczak and Viemeister, 2005; Moore et al., 2009; Wojtczak et al., 2011) suggest that modulation adaptation can occur after relatively brief stimulus exposures (i.e., on the order of hundreds of milliseconds), and thus adaptation could still have been a factor in the between-trial randomization condition. Therefore, we cannot rule out the possibility that adaptation to the masker modulation played a role in producing modulation IM in the main experiment.

Finally, we note that off-frequency listening in the modulation domain in the *F* condition may have been a factor in producing the positive threshold differences, and thus modulation IM, in the main experiment. The idea here is that, in the *F* condition, listeners could have attended to a modulation channel tuned to a frequency below the modulation frequency of the target for maskers that were above it (or vice versa) in order to improve the within-channel representation of the target. This strategy would have been difficult to deploy in the *R* condition, however, due to the randomization of the masker across trials, yielding poorer thresholds and modulation IM. While our results are insufficient to determine the extent to which off-frequency listening was a factor, it seems plausible, and therefore must be considered in future work in this area.

2.5. Summary and Conclusions

The purpose of this study was to determine if masker-modulation-frequency uncertainty in an otherwise typical modulation-masking experiment can have an adverse effect on target-modulation detection performance analogous to the adverse effect of masker-frequency uncertainty typically observed in studies of frequency IM. It did, and we took this finding as evidence for modulation IM. Certain aspects of the results lent credence to the hypothesis that modulation IM, by way of analogy to frequency IM, reflects an adverse effect of masker-modulation-frequency uncertainty on the ability to focus selective attention on a target modulation channel. Other aspects, however, were deemed equivocal, and alternative interpretations were considered based on (1) an increase in internal noise under conditions of uncertainty, (2) detection of the target modulation based on a change in subjective masker quality, (3) modulation adaptation, and (4) off-frequency listening in the modulation domain. Each of these interpretations appears plausible in accounting for certain aspects of the results and thus further work is required to fully understand the mechanisms responsible for producing modulation IM.

3. CUES TO REDUCE MODULATION INFORMATIONAL MASKING

3.1. Abstract

The detectability of target amplitude modulation (AM) can be reduced by masker AM in the same carrier-frequency region. It can be reduced even further, however, if the AM rate of the masker is uncertain [Conroy and Kidd, J. Acoust. Soc. Am. **149**, 3665-3673, (2021)]. This study examined the effectiveness of contextual cues in reducing this latter, uncertainty-related effect. Listeners were tasked with detecting fixed-rate target sinusoidal AM (SAM) in the presence of masker SAM applied simultaneously to the same broadband-noise carrier. A single-interval, two-alternative forced-choice detection procedure was used to measure sensitivity for the target SAM and masker AM-rate uncertainty was created by randomly selecting the AM rate of the masker SAM on each trial. Relative to an uncued condition, a pretrial cue to the masker SAM significantly improved sensitivity for the target; a cue to the target SAM, however, did not. The delay between the cue-interval offset and trial-interval onset did not affect the size of the masker-cue benefit, suggesting that adaptation of the masker was not responsible. A simple model of modulation masking captured some, but not all, trends in the psychophysical data, suggesting that a reduction of masker AM-rate uncertainty may have contributed to the masker-cue benefit.

3.2. Introduction

The detectability of target amplitude modulation (AM) can be reduced by masker AM in the same carrier-frequency region, an effect referred to as modulation masking (Bacon and Grantham, 1989; Houtgast, 1989). Modulation masking is tuned in the AM domain, meaning that the detectability of the target is poorest (i.e., masking is greatest) when the target and masker have similar, rather than dissimilar, AM spectra. As such, it has been suggested that modulation masking may be analogous to energetic masking (EM) in the audio-frequency domain insofar as it may reflect competition between the target and masker within relatively "peripheral" (i.e., peripheral relative to some more central processor; see Durlach et al., 2003a), rate-selective, neural channels. That is, EM occurs when a target and masker contain frequency components that fall within the passband of the same peripheral frequency channel(s) at the same or a similar time and the masker components dominate the output. By way of analogy, then, modulation masking may occur when a target and masker contain AM components that fall within the passband of the same, relatively "peripheral," AM channel(s) at the same or a similar time and the masker components dominate the output. In both cases, the detectability of the target—and thus masking—is determined primarily by the extent to which the detection-relevant target components (frequency or AM) are represented internally at the output of the "peripheral" channel(s) in question (hereafter, we drop the scare quotes).

In the audio-frequency domain, a distinction often is made between masking that is peripheral in origin and masking that is more central in origin. Peripheral masking is synonymous with EM, whereas central masking, and, in particular, that which is the result of masker-frequency uncertainty (i.e., uncertainty regarding a masker’s frequency spectrum), is associated with informational masking (IM; see Durlach et al., 2003a; Kidd et al., 2008). Recently, Conroy and Kidd (2021) suggested that a similar distinction between EM and IM may apply in the AM domain as well (see also Durlach et al., 2003a; Sheft and Yost, 2007). The basis for their suggestion was the finding that, in a masked AM-detection task characterized by a high degree of masker AM-rate uncertainty (i.e., uncertainty regarding a masker’s AM spectrum), modulation masking alone was insufficient to explain the listeners’ performance: target-AM-detection thresholds were higher (poorer) and psychometric functions (*d’* versus target-modulation depth) were shallower (evincing greater noise in the detection/decision process) when masker AM-rate uncertainty was high than when it was low, even though modulation masking (i.e., EM in the AM domain) was equated across the low- and high-uncertainty conditions. By way of analogy to IM in the audio-frequency domain, or “frequency IM” (cf. Durlach et al., 2003a), Conroy and Kidd suggested that this finding was consistent with a form of IM in the modulation domain, or “modulation IM.” (For a more detailed exposition of the parallels between frequency IM and modulation IM, as well as a more detailed definition of modulation IM itself, see Conroy and Kidd.) The purpose of this study was to extend that work and to determine if, in a similar, masked AM-detection task characterized by a high degree of masker AM-rate uncertainty, contextual cues to certain features of an otherwise highly uncertain stimulus AM spectrum could be used by listeners to reduce the effects of the uncertainty and thus to reduce the amount of modulation IM.

The ability to use contextual cues to reduce the effects of masker uncertainty and increase the detectability of a target sound has been demonstrated previously, although not, to our knowledge, for the specific case of masker AM-rate uncertainty. For the more studied case of masker-frequency uncertainty, for example, it has been shown repeatedly that contextual cues to the frequency spectrum of an otherwise highly uncertain, multitone masker can be used by listeners to increase the detectability of a target tone (or tone sequence) embedded in it relative to the detectability of that target tone (or tone sequence) in the absence of such cues (e.g., Richards and Neff, 2004; Richards et al., 2004; Kidd et al., 2011; Cao and Richards, 2012; Shen, 2017; Richards et al., 2021). Under similar "multitone-masking" conditions (cf. Neff and Green, 1987), a cue to the target-tone's frequency also can be beneficial and, notably, this is true whether the target tone's frequency is fixed and "certain" or random and uncertain on each trial (e.g., Richards and Neff, 2004). Presumably, therefore, the cues in both cases engage central mechanisms that work to counteract the effects of the masker uncertainty, although, in the case of a cue to the masker's frequency spectrum, it is also possible that more peripheral mechanisms are involved (e.g., adaptation of, or adaptation of the inhibition produced by, the masker frequency channels; see, e.g., Viemeister, 1980; Viemeister and Bacon, 1982). Richards et al. (2004), Kidd et al. (2011), Cao and Richards (2012), and Shen (2017) all have investigated this issue in considerable detail, and while their data suggest that both peripheral and central mechanisms likely contribute to the masker-cue benefit under frequency IM, they also suggest that when masker-frequency uncertainty is high, central mechanisms likely contribute more strongly (see also Byrne et al., 2013; Feng and Oxenham, 2015). Our question in this study was if, in an analogous, masked

AM-detection task characterized by a high degree of masker AM-rate uncertainty, contextual cues to either the target or masker AM rate could engage these or similar central mechanisms to reduce the effects of masker AM-rate uncertainty and thus to reduce the amount of modulation IM.

As a means to address this question, we employed a pretrial acoustic cuing paradigm very similar to the one that has been used previously to investigate cue effectiveness under multitone-masking conditions high in masker-frequency uncertainty (e.g., Richards and Neff, 2004; Richards et al., 2004; Cao and Richards, 2012; Shen, 2017; Richards et al., 2021). Listeners were tasked with detecting fixed-rate target sinusoidal AM (SAM) in the presence of masker SAM applied simultaneously to the same carrier (broadband noise). A single-interval, two-alternative forced-choice detection procedure was used to measure sensitivity for the target SAM and masker AM-rate uncertainty (i.e., modulation IM) was created by randomly selecting the AM rate of the masker SAM on each trial. In different conditions, each single-interval detection trial was either presented alone or was preceded by a pretrial acoustic cue to either the fixed-rate target SAM (target-cue condition) or the random-rate masker SAM (masker-cue condition). The effectiveness of the different cue types in reducing the effects of masker AM-rate uncertainty, and thus modulation IM, was determined by comparing the listeners' sensitivity for the target under cued and uncued conditions. We hypothesized that, as in the case of masker-frequency uncertainty, both cue types would be effective in reducing the effects of masker AM-rate uncertainty, and thus that both cue types would be effective in reducing modulation IM. As a general framework for interpreting cue effectiveness, we began with the premise that modulation IM, by way of analogy to frequency IM (cf. Lutfi, 1993; Neff et al., 1993; Wright and Saberi, 1999; Durlach et al., 2005), may reflect an adverse effect of masker AM-rate uncertainty on the ability to focus or maintain selective attention on a target-AM channel, an idea that received some qualified support in the study of Conroy and Kidd (2021). By this view, a benefit of either cue type in terms of reducing modulation IM should reflect an improvement in selective attention—relative to the uncued condition—about and/or to the target-AM channel. Indeed, under multitone-masking conditions, the benefit of a cue to a target tone's frequency and to a multitone-masker's frequency spectrum both have been interpreted along similar lines (e.g., Richards and Neff, 2004; Richards et al., 2004; Cao and Richards, 2012).

3.3. Methods

3.3.1. Listeners

Eleven listeners participated in this study, including the author (C.C.). Four listeners participated in Experiment 1 (three females, 20–31 years, mean = 24 years) and seven (different) listeners participated in Experiment 2 (five females, 19–25 years, mean = 21years).[6] All listeners had pure-tone air-conduction thresholds at or below 20 dB HL at octave frequencies from 250 to 4000 Hz in both ears. All listeners save for the author

---

[6] One additional observer was recruited for experiment 2 but was excluded from the final analysis for failing to reach a predetermined criterion level of performance of over 90% correct in an unmasked control condition included simply to ensure that the online observers (see main text) were engaged and performing the online task.

were naïve to the purposes of the study and received compensation for their participation. Recruitment and use of human subjects protocols were approved by the Boston University Charles River Campus Institutional Review Board.

3.3.2. Procedures

Data collection for Experiment 1 was completed in the laboratory. Stimuli were generated digitally at a sampling rate of 44100 Hz, routed through a 24-bit sound card (RME Digiface, Haimhausen, Germany), and presented monaurally to the listeners' left ears via a pair of circumaural earphones (Sennheiser, HD280 Pro, Wedemark, Germany). Listeners were tested individually while seated in a double-walled, sound-attenuated booth (Industrial Acoustics Company, North Aurora, IL), inside of which was a computer monitor, a computer keyboard, and a computer mouse. Stimulus generation and response collection was handled using custom scripts programmed in MATLAB (MathWorks, Natick, MA). Data collection for Experiment 2 was completed online. A MATLAB experiment was programmed and compiled as a web application and run on a Windows 2019 Virtual Machine with MATLAB Web App Server installed. The web application was open to the internet and the listeners accessed it via an internet browser. Stimuli were generated on the Virtual Machine on a trial-by-trial basis at a sampling rate of 44100 Hz, saved to the listeners' browser storage, and then presented as uncompressed WAV files monaurally to the listeners' left ears via their own computers' sound cards and their own earphones. The listeners interacted with the experiment via a graphical user interface (GUI) presented on the in-booth monitor (Experiment 1) or internet browser window (Experiment 2). The listeners' trial-by-trial responses were registered via mouse-click on labeled buttons displayed on the GUI.

In both experiments, a single-interval, two-alternative forced-choice detection procedure was used to measure sensitivity for target SAM (the target) applied with a 0.5 *a priori* probability on each trial to the full duration of a 500-ms broadband-noise carrier. In all conditions, masker SAM (the masker) was applied with a 1.0 *a priori* probability on each trial to the full duration of the same carrier. The AM rate of the target was always 32 Hz and thus was fixed and “certain” on each trial. The AM rate of the masker, by contrast, was selected at random from a wide range, a manipulation that was intended to produce large amounts of masker AM-rate uncertainty and thus large amounts of modulation IM (Conroy and Kidd, 2021). The procedure for selecting the AM rate of the masker on each trial was as follows: first, a collection of 20 possible masker rates, log-spaced between 4 and 256 Hz, was defined and the two rates nearest to the AM rate of the target (32 Hz) were removed to establish a protected zone (roughly +/-25% around 32 Hz). The protected zone was intended to reduce the relative influence of modulation masking in all conditions and thus to increase the relative influence of modulation IM. This left 18-possible, “base”-masker rates (nine below and nine above 32 Hz) from which to draw on each trial. On each trial, then, a base masker rate was selected at random with replacement from among the 18-possible base-masker rates and then an “actual”-masker rate was selected at random from a uniform distribution spanning a factor of +/-1.1 around the randomly selected base-rate value. This additional randomization/jitter stage was included to further increase the amount of uncertainty and reduce the probability of exact masker repetitions across subsequent trials. In all conditions of both experiments, the modulation depth of the masker (i.e., its modulation index, *m*) was fixed at 0.52. In Experiment 1, two target-modulation depths were tested: a "low" target depth ($m = 0.30$) and a "high" target depth ($m = 0.48$). In Experiment 2, only the low target depth ($m = 0.30$) was tested. On all trials of all conditions of both experiments, the starting phases of the target and/or masker were separately selected at random from a uniform distribution spanning 0 to 2π.

In both experiments, two types of condition were tested: cued and uncued. In the uncued conditions, the listeners heard a single sound on each trial (a "trial interval") which consisted of a broadband-noise carrier modulated by either the masker alone (masker-alone trials) or the target plus masker (target-plus-masker trials). The listeners' task was to indicate whether the target was present or absent. In the cued conditions, by contrast, the listeners heard two sounds on each trial (a "cue interval" followed by a trial interval) and made a judgement about whether the target was present in the second sound (i.e., the trial interval). In the cued conditions, the cue interval consisted of the same 500-ms broadband-noise carrier as the trial interval (i.e., the carrier was frozen across the cue and trial intervals of each trial) modulated by either the target alone (target-cue condition, Experiment 1) or the masker alone (masker-cue conditions, Experiments 1 and 2). The cue interval preceded the trial interval by an unfilled (i.e., "silent") inter-stimulus interval (ISI). In the cued conditions of Experiment 1, the ISI was 500 ms, as measured from the cue-interval offset to the trial-interval onset. In the cued condition of Experiment 2, the ISI was either 25 or 500 ms. On all cued trials of both experiments, the cue-interval stimulus was matched to the trial-interval stimulus in all relevant respects: as already noted, the carrier was identical across the cue and trial intervals, and so was the relevant modulator’s (i.e., target or masker) depth and phase. On masker-alone trials in the masker-cue conditions, for example, the cue- and trial-interval stimuli were identical. Following the application of modulation, all stimuli were ramped using 5-ms cosine-squared onset-offset ramps, filtered between 250 and 4000 Hz using an eighth-order Butterworth bandpass filter, and scaled to a fixed overall level. In Experiment 1, this level was 80 dB SPL. In Experiment 2, it was a “comfortable” overall level determined separately for each listener on the basis of a perceptual calibration routine performed prior to each online testing session, in which the listeners adjusted the level of an unmodulated broadband-noise (presented monaurally to the listeners’ left ears after filtering and ramping as in the experiment) to a “comfortable” overall level.

Data collection was blocked by condition as well as by target-modulation depth (Experiment 1) or ISI (Experiment 2). Each listener completed six blocks of 50 trials each for each condition-by-target-modulation-depth (Experiment 1) or condition-by-ISI (Experiment 2) combination. Correct answer feedback (correct/incorrect) was provided on all trials. Data collection for Experiment 1 was structured such that a given condition-by-target-modulation-depth combination was tested for three blocks in a row before proceeding to a different condition-by-target-modulation-depth combination. For each listener, data collection always began with three blocks of a random condition at the high target-modulation depth. Following these first three blocks, data collection proceeded such that the two remaining conditions were tested in a random order at the high target-modulation depth. Then, three blocks of each condition were tested at the low target-modulation depth in the same manner, with the condition order randomized anew. Once completed, this full sequence was repeated (with a different random permutation of condition order at each target-modulation depth) to arrive at the final data set for each listener. Data collection for Experiment 2 was structured such that the different experimental conditions (i.e., the uncued and masker-cue conditions) were tested in different experimental sessions. For the masker-cue session, the 25-ms and 500-ms ISI blocks were evenly distributed throughout the session, with the order in which the two ISIs were tested randomized for each two-block sequence. Four listeners completed the masker-cue session first and three completed the uncued session first. In addition to the experimental conditions, the Experiment 2 listeners also completed an unmasked control condition at the beginning of the uncued session. This condition was included simply to ensure that the listeners were engaged and performing the online task. In this unmasked control condition, the stimuli and task were identical to the uncued condition except that the modulation depth of the masker was set to zero. A predetermined performance criterion of at least 90% correct was required in this condition for a listener's data to be considered reliable and included in the final analysis. One listener failed to reach this criterion level of performance and thus was not included in the final analysis.

For each listener, an estimate of *d'* was obtained for each block of trials on the basis of the hits and false alarms registered in that block and *d'* was computed using the standard equation for single-interval tasks (see Equation 2.3). Extreme hit and false-alarm probabilities were adjusted prior to calculating *d'* using the "log-linear" rule (Hautus, 1995). The final estimate of *d'* for each listener for each condition-by-target-modulation-depth (Experiment 1) or condition-by-ISI (Experiment 2) combination was taken as the mean of the six *d'* estimates provided for each unique stimulus configuration.

3.3.3. Model of modulation masking

Modulation IM, by definition, is “nonenergetic” masking in the AM domain, with “energetic” masking in this context referring to modulation masking as defined in Section 3.2. It was important, therefore, to determine the extent to which modulation masking may have differed across the different cued and uncued conditions before drawing any strong conclusions regarding differences in modulation IM (i.e., cue-induced reductions of modulation IM). Towards this end, a simple, single-channel model of modulation masking was used to generate predictions for the different conditions and these predictions were compared to the psychophysical results. The logic was simple: appreciable differences between the model predictions and the psychophysical results should indicate that factors other than modulation masking, such as modulation IM, played a role in the listeners’ performance. More specifically, we reasoned that any cue benefits observed psychophysically not predicted by the model would be consistent with the suggestion of a cue-induced reduction of modulation IM.

The model was very similar to the envelope power spectrum model (EPSM) described by Ewert and Dau (2000). The most important difference between our model and the EPSM was that, in our model, the decision variable (DV) differed across conditions (cued versus uncued). In the uncued condition, the DV was the modulation power at the output of the target-AM channel, i.e., an AM channel tuned to 32 Hz, the AM rate of the target. This quantity was obtained by multiplying the magnitude spectrum of the ac-coupled envelope on each trial (i.e., the modulator that was applied to the carrier) by the squared magnitude transfer function of a second-order Butterworth bandpass filter with a center frequency of 32 Hz and a quality factor (Q; center frequency divided by bandwidth) of 1 and then integrating across AM rate (Ewert and Dau, 2000). This obtained quantity was then converted to dB to arrive at the DV. In the cued conditions, by contrast, the DV was the *difference* in modulation power at the output of the target-AM channel across the cue and trial intervals. This quantity was obtained by first separately estimating the modulation power at the output of the target-AM channel for the cue and trial intervals using the same second-order Butterworth bandpass filter, Q value, and integration procedure as in the uncued condition. The DV was obtained, however, by taking the difference in the modulation power estimates across the cue and trial intervals (trial interval−cue interval, signed in dB). For all conditions, a sensitivity index comparable to *d'*, d_a (Simpson and Fitter, 1973), was then computed via the following equation:

$$d_a = \frac{\mu_{T+M} - \mu_M}{\sqrt{(\sigma_{T+M}^2 + \sigma_M^2)/2}} \qquad (1)$$

where μ_{T+M} and μ_M were the means and σ_{T+M} and σ_M were the standard deviations of the DV on target-plus-masker and masker-alone trials, respectively.

3.4. Experiment 1

3.4.1. Results

The results of Experiment 1 are shown in Figure 3.1. The left panel shows the group-mean estimate of *d'* for each condition at both the low (white bars) and high (black bars) target-modulation depths, whereas the right panel shows the group-mean cue benefit (cued−uncued *d’*) for the two cued conditions at the same two target-modulation depths (white bars: low depth; black bars: high depth). The black error bars are standard errors of the mean (SEM), whereas the gray error bars are means of the standard errors computed for the individual listeners (i.e., the mean of the standard error across the six estimates of *d’* provided by each listener). Because we were relatively unconcerned in this study with absolute sensitivity and instead were concerned, primarily, with the extent to which the different cue types provided a detection benefit (i.e., yielded an increase in *d’*) relative to the uncued condition, we focus our discussion below on the cue benefit metric rather than on the absolute estimates of *d’*.

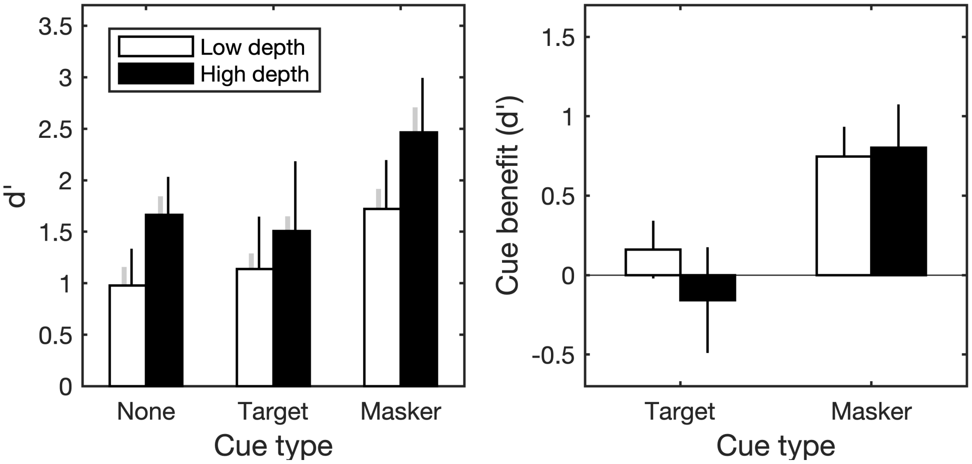

Figure 3.1. Left panel: Group-mean estimates of *d’* for each condition at both the low (white bars) and high (black bars) target-modulation depths. The black error bars give the standard error of the mean (SEM), whereas the gray error bars are the means of the standard errors computed for individual listeners. Right panel: Group-mean cue benefits (cued−uncued *d’*) for the two cued conditions at both the low (white bars) and high (gray bars) target-modulation depths. Error bars are SEM.

Masker cues provided a larger detection benefit than did target cues. Indeed, target cues tended not to provide a detection benefit at all. This was true at both the low and high target-modulation depths. At the low target depth, the group-mean masker-cue benefit (SEM) was 0.75 (0.19) and at the high target depth it was 0.80 (0.27). By contrast, the group-mean target-cue benefits were 0.16 (0.18) and -0.16 (0.33) at the low and high target-modulation depths, respectively. A two-way, repeated-measures analysis of variance (ANOVA) performed on the absolute estimates of *d'* indicated that the main effects of target-modulation depth [$F(1,3) = 58.505$, $p = .005$] and cue type [$F(2,6) = 8.402$, $p = .018$] both were significant, while the target-modulation-depth-by-cue-type interaction was not [$F(2,6) = 2.834$, $p = .136$]. Accordingly, Figure 3.2 shows the results of Experiment 1 averaged across the two target-modulation depths to better illustrate the effect of cue type. Clearly, masker cues provided a detection benefit, while target cues did not.

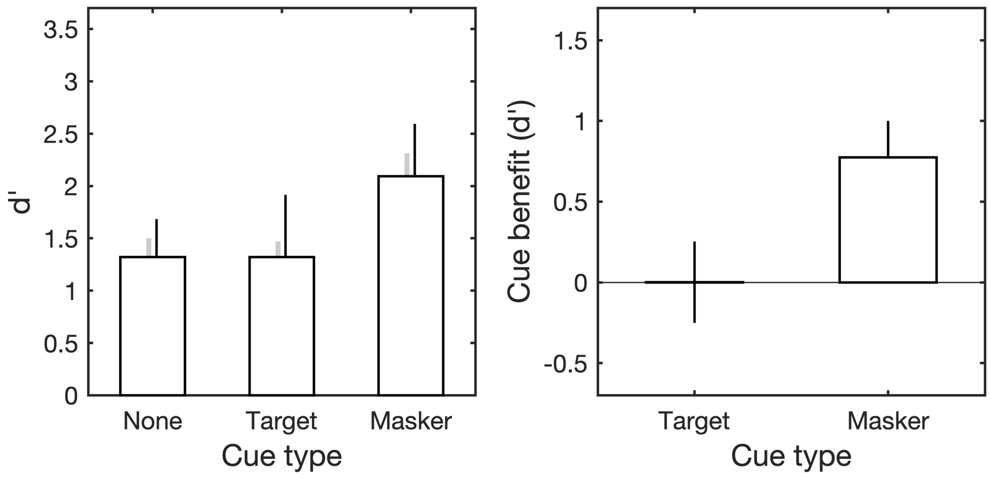

Figure 3.2. As in Figure 3.1 but with *d'* averaged across the two target-modulation depths prior to averaging across listeners.

Paired samples *t*-tests (single-sided) confirmed that while masker cues did indeed yield a significant improvement in *d'* relative to the uncued condition [$t(3) = 3.425$, $p = .021$], target cues did not [$t(3) = 0.005$, $p = .498$]. Averaged across the two target-modulation depths, the group-mean masker-cue benefit (SEM) was 0.77 (0.23); the group-mean target-cue benefit was 0.00 (0.25). Thus, the results of Experiment 1 were consistent with only one half of our initial hypothesis (i.e., that both cue types would be effective in reducing the effects of masker AM-rate uncertainty/modulation IM) and diverged from the results of previous studies on cue effectiveness under multitone-masking conditions, where a cue to a target-tone's frequency and to a multitone-masker's frequency spectrum both can provide substantial detection benefits relative to uncued conditions (e.g., Richards and Neff, 2004; Richards et al., 2004; Cao and Richards, 2012; Shen, 2017; Richards et al., 2021).

3.4.2. Model predictions

The complete lack of a target-cue benefit clearly demonstrates that the target cues were insufficient to yield a reduction of modulation IM. But what about the masker cues? That is, the detection benefit yielded by the masker cues could have reflected a number of factors, including, in particular, a reduction of modulation masking, rather than modulation IM. This would suggest that, contrary to our initial hypothesis, neither cue type was sufficient to reduce the effects of masker AM-rate uncertainty and, rather, that the masker cues provided a detection benefit relative to the uncued condition simply because they reduced the competition between the target and masker within one or more relatively peripheral, rate-selective neural channels. In an effort to rule out this possibility, we compared the psychophysical results to the model predictions for the different conditions tested in Experiment 1. As noted in Section 3.3.3, we reasoned that a substantial deviation between the predicted and observed masker-cue benefits would suggest the involvement of modulation IM.

The left panel of Figure 3.3 shows the model d_as for the three conditions tested in Experiment 1, whereas the right panel shows the model cue benefits (cued−uncued d_a) for the two cue types. White bars show the model d_as/cue benefits for the low target-modulation depth, whereas black bars show the model d_as/cue benefits for the high target modulation depth. To arrive at the model predictions shown in Figure 3.3, Experiment 1 was simulated on four model listeners (the same number of listeners as participated in the psychophysical experiment), each of whom completed 1000 trials (500 masker alone, 500 target plus masker) for each condition-by-target-modulation-depth combination. The model d_as/cue benefits shown in Figure 3.3 are group means across the four model listeners. Error bars in both panels are SEM.

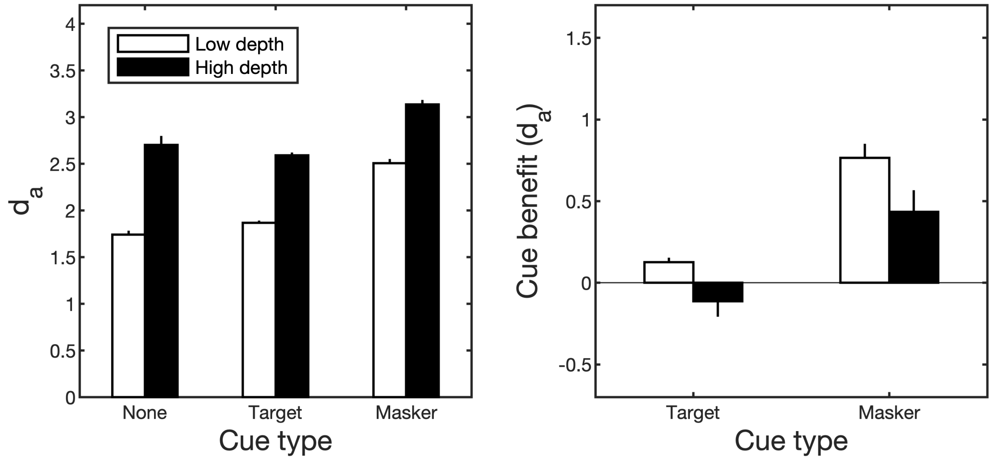

Figure 3.3. Model d_as (left panel) and cue benefits (cued−uncued d_a) for the different conditions tested in Experiment 1 plotted according to similar conventions as the psychophysical results shown in Figure 3.1. Model d_as/cue benefits are the means across four model listeners, each of whom completed 1000 trials for each stimulus configuration. Error bars in both panels are SEM.

On average, the model listeners were more sensitive than the psychophysical listeners overall (note the difference in scale of the y-axes between the left panels of Figures 3.1 and 3.3). This was unsurprising given that the model d_as were obtained using “out of the box” parameters (e.g., a Q value of 1) and without making any provision for the effects on d_a of the inherent modulations of the broadband-noise carrier (Dau et al., 1999), peripheral auditory processing, or, perhaps most importantly, internal noise (Ewert and Dau, 2004). Nevertheless, the trends in the model d_as across conditions, and thus the pattern of cue benefits across the two cue types, closely resembled the psychophysical results: specifically, the model predicted a masker-cue benefit but effectively no target-cue benefit at both target-modulation depths. Notably, however, unlike in the psychophysical results, the model predicted an appreciable difference in the size of the masker-cue benefit at the two target-modulation depths, with a larger predicted masker-cue benefit at the low target-modulation depth than at the high target-modulation depth.

At the low target-modulation depth, the predicted group-mean masker-cue benefit (SEM) was 0.76 (0.09), while at the high target depth it was 0.43 (0.13). Thus, the predicted masker-cue benefit at the low target depth was very similar in magnitude to psychophysical cue benefit but, at the high target depth, it was smaller than that observed psychophysically. Nevertheless, the obvious correspondence between the model predictions and the psychophysical results suggests that the masker cues may have provided a detection benefit via a reduction of modulation masking (i.e., EM in the AM domain), rather than modulation IM.[7]

3.4.3. Discussion

A deviation of the model predictions from the psychophysical results would have been sufficient to rule out an interpretation of the masker-cue benefit in terms of modulation masking alone, or, at least, modulation masking as formalized in our simple, single-channel model. The opposite, however, is not true, and there are at least two other interpretations of the masker-cue benefit that cannot be ruled out at this juncture: one based on a reduction of modulation IM, as hypothesized, and another based on adaptation of masking (Viemeister, 1980) in the AM domain (i.e., a different mechanism by which

---

[7] It is worth reflecting on how “modulation masking” is reduced in the model and whether the reduction of modulation masking yielded by the masker cues is consistent with a reduction of EM as traditionally defined. That is, in the model, a different DV is employed under cued and uncued conditions. Obviously, for human listeners, shifting from one DV to another depending on stimulus context would involve central mechanisms and so, in a certain sense, the masker-cue benefit predicted by the model—which ultimately reflects a centrally mediated incorporation of contextual AM information during the calculation of the DV—can be said to reflect a reduction of modulation IM, rather than modulation masking. In this chapter, however, we take the opposite view simply because, in the model, the detectability of the target *in all conditions* (cued and uncued) is determined solely by the extent to which the different masker modulators stimulate the target AM channel. Admittedly this is something of a conceptual gray area.

modulation masking may have been reduced in the masker-cue condition relative to the uncued condition). In this section, we consider each of these interpretations in turn.

Conroy and Kidd (2021) found some qualified support for the notion that modulation IM, by way of analogy to frequency IM under multitone-masking conditions (cf. Lutfi, 1993; Neff et al., 1993; Wright and Saberi, 1999; Durlach et al., 2005), may reflect an adverse effect of masker AM-rate uncertainty on the ability to focus or maintain selective attention on a target-AM channel. By this view, the masker-cue benefit should have corresponded to an improvement in the listeners' attention, relative to the uncued condition, about and/or to the target-AM channel. At first blush, this interpretation seems implausible: if the masker cues promoted the focusing and/or maintenance of the listeners' attention about and/or to the target-AM channel, why did the target cues fail to do the same? Surely, a direct cue to the target channel in the form of direct, pretrial acoustic stimulation would have been highly advantageous to attentional focusing. Recall, however, that the AM rate of the target was fixed and therefore "certain" on each trial (i.e., there was no experimentally imposed target AM-rate uncertainty). As such, it is possible that the listeners did not need such a pretrial reminder in order to attend to the target channel and instead experienced more masking (i.e., modulation IM) in the uncued and target-cue conditions than in the masker-cue condition because, in the uncued and target-cue conditions, the uncertain masker drew their attention away from the target-AM channel on some trials (for a similar account of frequency IM under multitone-masking conditions, see Lutfi, 1993; Oh and Lutfi, 1998; Lutfi et al., 2003b). One way to avoid such distractions of attention/target-masker confusions would have been to suppress, via central mechanisms, the activity within the AM channel most sensitive to the masker on each trial (i.e., to employ a “Listener-Min” strategy; cf. Durlach et al., 2003a). Because this would have been possible with the masker cues, but not the target cues, a masker-but-not-target-cue benefit would be expected. In other words, the pattern of cue benefits reported here.

Alternatively, it is possible that adaptation of masking in the AM domain played a role in producing the masker-cue benefit. The idea here is that, on each trial of the masker-cue condition, the masker cue served to adapt the response of the AM channel most sensitive to the masker, increasing the relative response strength of the target channel in turn. Indeed, as noted in Section 3.2, it is likely that adaptation of masking and/or of inhibition (Viemeister and Bacon, 1982) in the audio-frequency domain has played at least some role in producing the masker-cue benefits that have been observed under frequency IM (Richards et al., 2004; Kidd et al., 2011; Cao and Richards, 2012; Byrne et al., 2013; Feng and Oxenham, 2015; Shen, 2017) and this, taken together with the suggestion of within-AM-channel adaptation effects coming from studies of both AM forward masking (Wojtczak and Viemeister, 2005) and AM adaptation (Tansley and Suffield, 1983; Richards et al., 1997; Wojtczak and Viemeister, 2003) under low-uncertainty conditions, suggests that the masker-cue benefit reported here may have included at least some influence of AM adaptation.

While plausible, certain aspects of Experiment 1 point away from an interpretation of the masker-cue benefit in terms on adaptation of masking in the AM domain, including, for example, the relatively long ISI that was used (500 ms). The results of AM forward masking experiments (the most relevant for our purposes given the duration of “adaptor” exposures in the current experiment) have demonstrated that the decay of AM forward masking (i.e., the recovery from AM adaptation, assuming that AM adaptation contributes to the AM forward masking effect) is exponential: forward-masked AM-detection thresholds improve rapidly over masker-target gap durations of 0-300 ms (Wojtczak and Viemeister, 2005; Wojtczak et al., 2011). Residual amounts of AM forward masking have been reported for masker-target gap durations as long as 510 ms (Wojtczak and Viemeister, 2005; Wojtczak et al., 2011), however, suggesting that, in the current experiment, recovery from AM adaptation could have been incomplete by trial-interval onset. Another reason one might doubt an interpretation of the masker-cue benefit in terms of adaptation of masking in the AM domain is the relatively low masker-modulation depth that was used ($m = 0.52$). Most previous studies of AM forward masking and AM adaptation, for example, have used fully modulated maskers ($m = 1$). In theory, adaptation should decrease with masker-modulation depth. Füllgrabe et al. (2021), however, found that masker modulators with depths of $m = 1.0$ and $m = 0.25$ produced comparable amounts of AM forward masking at short masker-target gaps (< 200 ms). It is possible, therefore, that the relatively low masker-modulation depths used here were sufficient to yield at least some adaptation of the masker-sensitive AM channel(s). Perhaps the most compelling reason for doubt regarding an adaptation-of-masking-based interpretation of the masker-cue benefit, however, is that the target cues failed to elicit a *negative* target-cue benefit, which would be expected due to adaptation of the target-AM channel by the target cues. Moreover, residual adaptation of the target channel due to the repeated presentation of the target cues (and, indeed, the target itself) across trials would be expected to exacerbate this effect, yielding larger, albeit negative, target-cue benefit than masker-cue benefit. Obviously, this is not what we found.

One place where an adaptation-of-masking-based interpretation of the masker-cue benefit differs from both the model of modulation masking and an interpretation based on a reduction of modulation IM via improved attentional focusing is in the predictions that it makes regarding the effect of ISI. More specifically, an adaptation-of-masking-based interpretation of the masker-cue benefit predicts a decrease in the size of the masker-cue benefit as a function of ISI due to the recovery from adaptation within the masker-sensitive channels, whereas the other two interpretations do not. Experiment 2, therefore, examined the effect of ISI on the size of the masker-cue benefit in an attempt to determine the extent to which adaptation of masking in the AM domain played a role in producing the masker-cue benefit observed in Experiment 1.

3.5. Experiment 2

3.5.1. Results

The results of Experiment 2 are shown in Figure 3.4.[8] The left panel shows the group-mean estimate of *d'* for the uncued condition, and for the masker-cue condition at the two ISIs (white bar: 25-ms ISI; black bar: 500-ms ISI). The right panel shows the

---

[8] Two observers completed the two experimental conditions from experiment 2 in the laboratory as well as online to ensure the reliability of the online *d'* estimates. The agreement between the online and laboratory estimates of *d'* was good (r = .91, p = .012), suggesting that the online-testing procedures yielded reliable estimates of *d'* for the stimuli and conditions tested here, despite yielding an increase in the variability of *d'* estimates across blocks (compare the size of the gray error bars in Figure 3.1 to those in Figure 3.4).

masker-cue benefit at the two ISIs (white bar: 25-ms ISI; black bar: 500-ms ISI). The black error bars in both panels are SEM, whereas the gray error bars in the left panel are means of the standard errors computed for individual listeners.

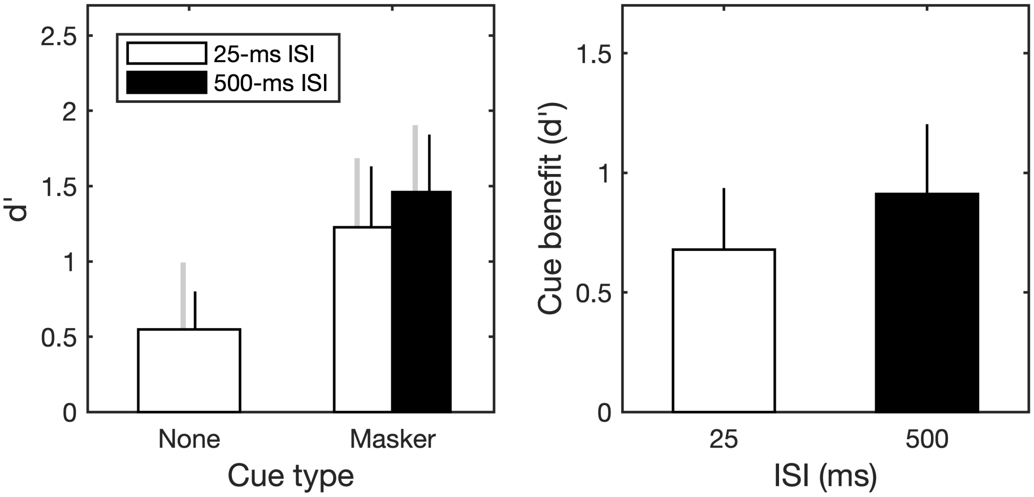

Figure 3.4. Group-mean *d'* (left panel) and masker-cue benefit (right panel) for the different conditions tested in Experiment 2. For the masker-cue condition, white bars give the group-mean *d'*/masker-cue benefit at the 25-ms inter-stimulus interval (ISI), while black bars give the group-mean *d'*/masker-cue benefit at the 500-ms ISI. Otherwise, conventions are as in Figures 3.1 and 3.2.

Masker cues provided a detection benefit at both ISIs, replicating the masker-cue benefit observed in Experiment 1. The size of the masker-cue benefit, however, tended not to differ appreciably between the two ISIs: for the 25-ms ISI, the group-mean masker-cue benefit (SEM) was 0.68 (0.26); for the 500-ms ISI, it was 0.91 (0.29). A paired samples *t*-test (two-sided) failed to reveal a significant difference in *d'* between the two ISIs [$t(6) = 1.507$, $p = .182$]. An additional *t*-test (single-sided, performed on the *d's* after averaging across the two ISIs), however, confirmed that the masker cues did indeed yield a significant improvement in *d'* relative to the uncued condition [$t(6) = 3.004$, $p =$

.012]. Thus, the results of Experiment 2 were inconsistent with an interpretation of the masker-cue benefit in terms of adaptation of masking in the AM domain.

3.5.2. Discussion

Experiment 2 was intended to test the hypothesis that the masker-cue benefit observed in Experiment 1 (and replicated in Experiment 2) was the consequence of adaptation of masking in the AM domain, rather than a reduction of masker AM-rate uncertainty/modulation IM or modulation masking (as formalized in the single-channel model outlined in Section 3.3.3). An adaptation-of-masking-based interpretation of the masker-cue benefit predicts that the size of the masker-cue benefit should depend on ISI due to the recovery from AM adaptation within the masker-sensitive channel(s). Experiment 2 tested this prediction by measuring the effectiveness of the masker cues at two ISIs—one “short” and one “long”—with the expectation that, if adaptation of masking was responsible for the masker-cue benefit, the size of the effect would be greater at the shorter ISI. The results of Experiment 2 did not conform to this expectation: at both ISIs, masker cues provided a detection benefit as they did in Experiment 1, but the size of the masker-cue benefit did not differ significantly between the two ISIs. This suggests that adaptation of masking in the AM domain was not an important factor in experiments 1 or 2.

As noted in Section 3.2 and 3.4.3, there is some evidence to suggest that adaptation of masking and/or *of inhibition* in the audio-frequency domain may contribute to the masker-cue benefits that have been observed under multitone-masking conditions. Translated into the AM domain, the idea with adaptation of inhibition is that, rather than provide a detection benefit by reducing the relative response strength of the masker-sensitive AM channel(s), the masker cues provided a detection benefit because they adapted the inhibition of the target channel produced by the masker channel(s). Because we know of no psychophysical evidence of the kind of across-AM-channel inhibitory mechanisms that would be required to support such an interpretation, we have focused our discussion thus far on adaptation of masking, rather than inhibition. Nevertheless, it is certainly possible that such inhibitory mechanisms exist and so, considering only the results of Experiment 1, an adaptation-of-inhibition-based interpretation of the masker-cue benefit would be reasonable. An adaptation-of-inhibition-based interpretation, however, would make the same prediction regarding the effect of ISI as does adaptation of masking and so the results of Experiment 2 are inconsistent with an interpretation based on adaptation of inhibition as well.

3.6. General discussion

Modulation IM is a type of masking that occurs when a listener is uncertain about a masker's AM spectrum (Conroy and Kidd, 2021). In the current study, such uncertainty was created by randomly varying the AM rate of a masker from trial-to-trial in a single-interval, masked AM-detection task. The AM rate of the target was always fixed. The question that was addressed was whether, under such conditions, contextual cues to either the fixed-rate target or the random-rate masker (SAM in both cases) could be used by listeners to reduce the effects of the uncertainty and thus to reduce the amount of modulation IM. On the basis of previous findings on cue effectiveness under frequency

IM (i.e., multitone-masking conditions high in masker-frequency uncertainty), we hypothesized that both cue types would be effective in reducing modulation IM. Contrary to this hypothesis, however, we found that while a pretrial cue to the otherwise highly uncertain masker AM rate significantly improved the detectability of the target SAM (Experiments 1 and 2), a pretrial cue to the target SAM did not (Experiment 1). In seeking to account for this pattern of results, three main interpretations were considered: (1) that the masker cues, but not the target cues, yielded a reduction of modulation masking at the output of the target-AM channel, an interpretation that was formalized in a simple, single-channel model based on the EPSM of Ewert and Dau (2000); (2) that the masker cues, but not the target cues, yielded a reduction of modulation IM, as hypothesized, because they helped to prevent distractions of attention away from the target-AM channel on some trials; and (3) that adaptation of masking (Viemeister, 1980) in the AM domain played a role in producing the masker-cue benefit. Experiment 2 demonstrated that the size of the masker-cue benefit did not depend on the duration of the delay between the cue-interval offset and the trial-interval onset (i.e., the ISI) as it should if adaptation of masking contributed appreciably to the effect. Thus, we ruled out this interpretation. Distinguishing between the other two interpretations, however, was more difficult: while the model of modulation masking captured the main trends in the psychophysical data (i.e., the benefit of masker cues but not target cues, and the overall size of the masker-cue benefit), it did not capture them all (e.g., the model predicted an interaction between the cue type and target-modulation depth that was not apparent in the psychophysical results and it was far too sensitive overall). More generally, the good model predictions were insufficient to rule out an interpretation of the masker-cue benefit based on attentional mechanisms because all aspects of the psychophysical data were consistent with an involvement of such mechanisms.

One way to gain some clarity on the issue of whether the masker cues provided a detection benefit primarily via a reduction of modulation masking or modulation IM is by examining the listeners' sensitivity for the target SAM as a function of the AM rate of the masker SAM (i.e., their tuning curves) under both cued and uncued conditions. If the tuning curves for either condition deviated substantially from the shape that would be expected on the basis of modulation masking alone, it can be argued that factors other modulation masking, such as masker uncertainty/modulation IM, influenced the listeners' performance in some way. More specifically, in the model, modulation masking occurs when there is competition between the target and masker within a single AM channel. This channel is modeled as a second-order Butterworth bandpass filter with a center frequency of 32 Hz and a Q value of 1. If modulation masking was the primary factor responsible for masking under both cued and uncued conditions, the listeners' tuning curves should largely trace the shape of the putative target channel: at the most general level, therefore, the model predicts a monotonic decrease in the listeners' sensitivity with increasing target-masker similarity due to increasing competition between the target and masker within the single, attended-to target channel.

To examine this prediction in more detail, we constructed tuning curves for each listener for both the uncued and masker-cue conditions in the following manner: first, the raw, trial-by-trial response data for each condition were pooled across the two target-modulation depths or ISIs and then sorted by masker rate (i.e., the AM rate of the masker on each trial). Then, an estimate of *d'* was obtained for each masker rate on the basis of the hits and false alarms registered at that rate. The log-linear rule was applied where necessary. Recall that the “actual”-masker rate that was tested on each trial was determined by first selecting a “base”-masker rate from one of 18 possible rates and then randomly selecting the “actual”-masker rate from a uniform distribution spanning a factor of +/-1.1 around this base-rate value. In calculating the masker-specific estimates of *d'*, the actual-masker rates were “rebinned” according to the 18-possible base-masker rates so as to increase the total number of trials that contributed to each estimate. Thus, while more than 18 masker rates were tested, only 18 estimates of *d'* were obtained for each listener and condition.

The left panel of Figure 3.5 shows the group-mean tuning curves for both the uncued (black bars) and masker-cue (white bars) conditions. Each *d'* estimate plotted in the left panel of Figure 3.5 represents the group-mean estimate of *d'* across the 11 listeners who participated in this study (i.e., all listeners from both experiments). Error bars are SEM. Note that, in the left panel of Figure 3.5 the black bar associated with each masker rate is plotted on top of the white bar. As such, a visible white bar at a given masker rate indicates a masker-cue benefit at that rate. The right panel of Figure 3.5 plots this directly, i.e., the group-mean masker-cue benefit (cued minus uncued *d'*) for each masker rate. Error bars in both panels are SEM.

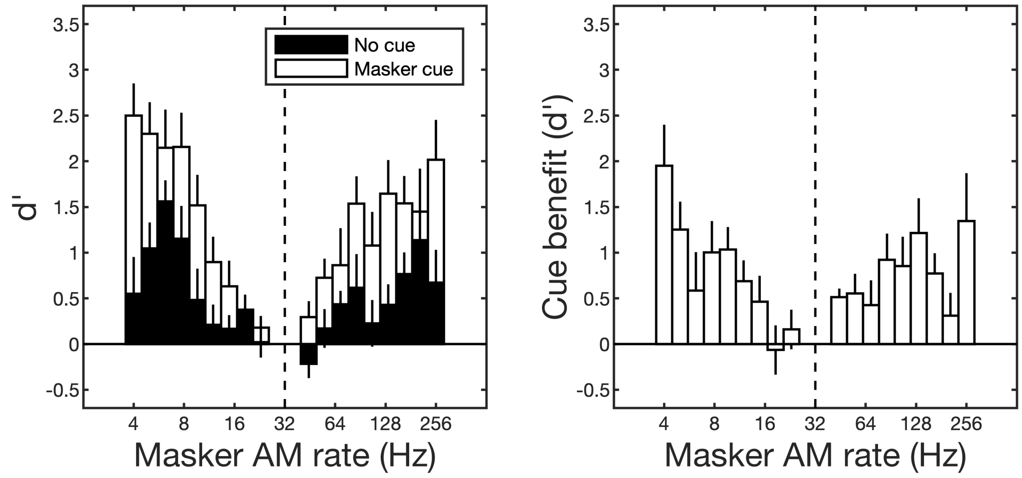

Figure 3.5. Group-mean *d'*s (left panel) and masker-cue benefits (right panel) plotted as a function of the amplitude-modulation (AM) rate of the masker on each trial. In the left panel, black bars give the *d'*s for the uncued condition, whereas the white bars (plotted behind the black bars) give the *d'*s for the masker-cue condition. The vertical dashed line in both panels marks 32 Hz, i.e., the AM rate of the target. Hence, the gap between bars in the region of the vertical dashed line demarcates the protected zone. Error bars are SEM.

There are two main aspects of the tuning curves shown in the left panel of Figure 3.5 that warrant mention. First is the tuning itself: the detectability of the target was poorest, and thus masking was greatest, when the target and masker had similar, rather than dissimilar, AM rates. This was clearly true in the masker-cue condition and, to a lesser extent, in the uncued condition. Such tuning is consistent with the suggestion that modulation masking played a role in both conditions. The second aspect of the tuning curves that warrants mention, however, is that there were marked differences in the shapes of the tuning curves across conditions. The uncued tuning curve, for example, was far from smooth in shape, and exhibited nonmonotonicities on both sides of the protected zone. This suggests that factors *other* than modulation masking, such as modulation IM, contributed to the overall amount of masking in the uncued condition. The masker-cue tuning curve, by contrast, showed no substantial nonmonotonicities on either side of the protected zone, suggesting that modulation masking may have been the primary factor contributing to masking in this condition. Taken together, then, the relative shapes of the uncued and masker-cue tuning curves suggest that at least some of the detection benefit yielded by the masker cues may have reflected a reduction of modulation IM, rather than modulation masking.

Recall that the model of modulation masking described in Section 3.3.3 largely captured the listeners' performance *at the condition level* (Figure 3.3). Without making any modifications to the model, however, it produces *masker-specific* estimates of d_a (i.e., tuning curves) that are dramatically greater than the psychophysical estimates of d' shown in Figure 3.5. The reason for this is that, as noted in Section 3.3.3, the model does not include an internal noise parameter, nor does it include the influence of the trial-by-trial variability in the stimulus envelopes produced by carrier. While we found that the condition-level estimates of d_a were high relative to the corresponding psychophysical estimates of d' (compare Figures 3.1 and 3.3), the two measures of sensitivity nevertheless were close—much closer than the masker-specific estimates of d_a and d' (not shown). This is because there was variability in the DVs that contributed to the condition-level estimates of d_a associated with the trial-by-trial variability in the masker AM rate. For the masker-specific estimates of d_a this variability was eliminated; hence, the d_as were enormous. One way to adjust the model to produce somewhat reasonable masker-specific estimates of d_a would be to introduce a central internal noise parameter (cf. Durlach and Braida, 1969) following the calculation of the DV, which could be conceptualized as a general uncertainty term related to criterion setting, target-memory decay, etc. If the variance of this noise was free to vary across conditions, a larger value in the masker-cue condition than in the uncued condition could be interpreted in terms of a masker-cue-induced reduction of uncertainty—that is, masker-cue-induced reduction of modulation IM. Indeed, informal model adjustments suggest that this could be a viable avenue to explore. An extensive model-fitting exercise, however, is outside the scope of the current study and so we leave it for future work.

One consequence of the nonmonotonicities in the uncued tuning curve shown in the left panel of Figure 3.5, taken together with the relatively smooth shape of the masker-cue tuning curve shown in the same panel, is that there were differences in the size of the masker-cue benefit at different masker rates. In particular, the largest group-mean masker-cue benefit occurred for those maskers that were most dissimilar from the target. Such differences are consistent with the notion that the masker cues provided a detection benefit, at least in part, because they enabled the listeners to avoid distractions of attention away from the target-AM channel on some trials. That is, assuming that modulation masking was a factor in both the uncued and masker-cue conditions, distractions of attention away from the target channel likely would have had the greatest relative influence on the overall amount of masking for those maskers that produced the smallest amounts of modulation masking. Thus, the relative shapes of the uncued and masker-cue tuning curves, and the pattern of masker-cue benefits across masker rates, are consistent with the suggestion that modulation IM in the current study may have reflected distractions of attention away from the target-AM channel on some trials, and that masker cues can reduce such distractions.

The masker-cue benefits reported in this study are reminiscent of Bacon and Grantham's (1992) results on fringe effects in modulation masking under low-uncertainty conditions. In that study, Bacon and Grantham measured the detectability of target SAM applied to a broadband-noise carrier in the presence of masker SAM applied to a second broadband-noise carrier. The carriers were continuous throughout each block of two-interval, two-alternative forced-choice trials while the target and/or masker SAM were applied during discrete observation intervals. They found that if the masker SAM was applied to its carrier for a period either before, after, or both before and after each observation interval, the detectability of the target SAM improved relative to when the masker-carrier fringes were left unmodulated. In all conditions, the rates of both the target and masker SAM were fixed across trials within each block and so, presumably, masker AM-rate uncertainty was low. Thus, the masker-fringe benefits presumably were unrelated to modulation IM. Bacon and Grantham proposed a number of potential mechanisms to explain their findings, including, in particular, those related to perceptual grouping. The idea was that, in the absence of a masker-modulator fringes, the target and masker SAM were perceptually grouped together due to their simultaneous onsets and offsets. The onset/offset asynchrony in the masker-fringe conditions, by contrast, yielded a situation in which the target SAM was perceptually segregated from the masker SAM and thus was detected more easily. A similar interpretation could be applied to our results, i.e., the masker-cue benefit could be interpreted as reflecting a reduction of modulation IM subsequent to the perceptual segregation of the target and masker (cf. Kidd et al., 1994; Neff, 1995). Conversely, it is possible that the target and masker were perceptually fused given the presence of strong low-level grouping cues (e.g., the fact that the target and masker shared the same carrier and had simultaneous onsets and offsets) and, as such, that the trial interval stimuli were discriminated from the cue interval stimuli on the basis of their higher-order representations (cf. Conroy and Kidd, 2021). It is worth noting, however, that this account predicts a target cue benefit as well.

3.7. Summary and conclusions

The purpose of this study was to examine the extent to which modulation IM could be reduced through contextual cues to either the target or masker SAM in a single-interval, two-alternative forced-choice masked AM-detection task characterized by a high degree of masker AM-rate uncertainty. It was found that, relative to an uncued condition, pretrial cues to the otherwise highly uncertain masker SAM yielded significant increases in target detectability; cues to the target SAM, by contrast, did not. A simple, single-channel model of modulation masking yielded predictions that were consistent with this general pattern of cue benefits, suggesting that a reduction of modulation masking (i.e., EM in the AM domain) may have contributed to the masker-cue benefit. Certain aspects of the psychophysical results, however, in particular the relative shape of the listeners' tuning curves under cued and uncued conditions, suggested that a reduction of modulation IM may have contributed as well. Potential mechanisms of this reduction include improvements in selective attention about and/or to the target-AM channel under cued versus uncued conditions, a reduction of internal noise (i.e., an overall decrease in uncertainty), and/or perceptual segregation of the target SAM from the masker SAM.

Further work is required to elucidate the relative contributions of these different mechanisms to the masker-cue benefits reported here.

4. PERCEPTUAL SEGREGATION OF MODULATION COMPONENTS UNDER MODULATION INFORMATIONAL MASKING

4.1. Abstract

The “multiple-bursts same, multiple-bursts different” (MBS-MBD) paradigm [Kidd et al., J. Acoust. Soc. Am. **95**, 3475-3480 (1994)] has been used extensively to examine the perceptual segregation of competing frequency components under conditions of masker-frequency uncertainty (informational masking or IM). In this study, we adapted the MBS-MBD paradigm to the amplitude-modulation (AM) domain to examine the perceptual segregation of competing AM components (peaks in the AM spectrum) under conditions of masker AM-rate uncertainty (“modulation IM”). Listeners were tasked with detecting a target comprising four sequential “bursts” of equal-rate, sinusoidal AM (i.e., a “coherent” sequence of AM) in the presence of a multiburst, multicomponent masker modulator applied to the same carrier. The AM components comprising the masker were either coherent (MBS condition) or incoherent (MBD condition) across bursts. The detectability of the target improved under MBD re MBS conditions, suggesting that the relative coherence of the competing AM components may have influenced the listeners’ performance. A “single-burst” detection strategy, however, could not be ruled out, indicating that sensitivity to relative coherence *per se* may *not* have been necessary to obtain an MBD re MBS benefit. The results highlight important similarities and differences between IM arising from masker uncertainty along different acoustic dimensions.

4.2. Introduction

The detectability of a pure-tone target flanked by a small number of other, spectrally remote pure tones of equal duration (collectively, masker) can be quite poor if the frequencies of the flankers are varied randomly from presentation-to-presentation, a finding often attributed to informational masking (IM; cf. Neff and Green, 1987). On any single stimulus presentation, the small number and spectral sparsity of the flankers is such that peripheral interactions resulting in energetic masking (EM) presumably are modest, yet the randomization of the flankers across stimulus presentations can induce uncertainty in the listener (in this case, masker-frequency uncertainty, or uncertainty regarding the masker's frequency spectrum), and such uncertainty can produce IM (cf. Durlach et al., 2003a; Kidd et al., 2008). One interpretation of IM in this context is that it occurs because the target tone, when present, is erroneously grouped together perceptually with the flankers, obscuring its presence, and producing IM. Indeed, under otherwise similar conditions, stimulus manipulations intended to yield a perceptual segregation of the target tone can yield substantial reductions of the amount of IM (e.g., Kidd et al., 1994; Neff, 1995; Durlach et al., 2003b). Presumably, such manipulations make it easier for some listeners to focus selective attention on the target, seeing as it is a distinct perceptual object segregated from the masker (cf. Shinn-Cunningham, 2008).

A particularly compelling demonstration of how perceptual grouping and segregation mechanisms can produce or reduce IM was reported by Kidd et al. (1994), who found that if multiple masker "bursts" comprising randomly selected flankers were presented in close temporal proximity over the course of each stimulus presentation, the detectability of a target tone (gated on and off with each burst) could improve dramatically relative to the single-burst (SB) random-flanker condition described above. Kidd et al. suggested that the release from IM provided by this “multiple-bursts-different” (MBD) manipulation was related to perceptual grouping and segregation mechanisms sensitive to “coherence,” i.e., the rapid, repeated presentation of a sound element or auditory feature across time (see also Kidd et al., 1995; Huang and Richards, 2006). That is, the coherent sequence of target tones, Kidd et al. argued, were grouped together perceptually into a single auditory object, which was perceived as segregated from the incoherent masker tones. Indeed, consistent with this interpretation, Kidd et al. also found that repeating the same collection of flankers across all bursts of each sequence (again, with the target tone gated on and off with each burst) provided no such improvement in performance relative to the SB condition. Presumably, performance remained poor for this “multiple-bursts-same” (MBS) manipulation because both the target and flankers formed coherent-frequency sequences. As such, they were grouped together perceptually into a single auditory object, making selective attention to the target much more difficult than it was when segregated (cf. Shinn-Cunningham, 2008).

Since Kidd et al. (1994) published their initial study, a number of others have used the same or a similar MBS-MBD-type paradigm to examine the perceptual grouping and segregation of competing frequency components under conditions of masker-frequency uncertainty (IM in the frequency domain, or “frequency IM”; cf., Durlach et al., 2003a). And, in general, the findings from those studies are consistent with Kidd et al.’s initial interpretation: that is, coherent frequency components tend to be grouped together perceptually into a single auditory object, which is perceived as segregated from incoherent components (e.g., Kidd et al., 1994; Kidd et al., 1995; Oxenham et al., 2003; Kidd et al., 2003; Durlach et al., 2003b; Richards and Tang, 2006; Huang and Richards, 2006, 2008; Micheyl et al., 2007; Micheyl et al., 2013; Teki et al., 2013). Indeed, the strength of such *relative* coherence as a grouping and segregation cue is so strong in such contexts, that some authors have argued that grouping and segregation based on relative coherence is a general perceptual principle that applies across feature dimensions (cf. Shamma et al., 2011), i.e., across dimensions along which separate auditory features are coded separately in the auditory system. This suggests, in turn, that perceptual grouping and segregation based on relative coherence—and the selective attention benefits this process provides in an acoustic mixture—may be one way that the auditory system copes with IM arising from masker uncertainty along dimensions other than frequency, such as space (Fan et al., 2008) or amplitude-modulation (AM) rate (Conroy and Kidd, 2021). To our knowledge, however, the frequency dimension is the only one in which relative-coherence-based grouping and segregation mechanisms have been examined explicitly in psychophysical experiments concerned with IM.

In the experiments reported in this chapter, we adapted the MBS-MBD paradigm to the AM domain to examine the perceptual grouping and segregation of competing AM components (peaks in the AM spectrum) under conditions of masker AM-rate uncertainty (IM in the modulation domain, or "modulation IM"; cf. Conroy and Kidd, 2021). Of primary interest was whether the relative-coherence-based grouping and segregation mechanisms revealed by previous MBS-MBD-type experiments operate in the AM domain as well as in the frequency domain, and thus whether perceptual grouping and segregation based on relative coherence can facilitate selective attention to, and information extraction from, particular regions of an otherwise highly uncertain stimulus AM spectrum. *A priori*, we had reason to think that it might. Specifically, there is both psychophysical (e.g., Bacon and Grantham, 1989; Houtgast, 1989) and neural (Langner and Schreiner, 1988) evidence of frequency selectivity in the AM domain, meaning that the individual components of a complex envelope may be coded separately in the auditory system (i.e., AM channels; cf. Dau et al., 1997; Ewert and Dau, 2000). A coherence-sensitive mechanism that operates across feature dimensions, then, should be sensitive to the relative coherence of competing AM components so long as the AM components in question are coded via separate mechanisms in the auditory system.

Our AM-domain version of the MBS-MBD experiment was very similar, conceptually, to the version described by Kidd et al. (1994). Listeners were tasked with detecting a target comprising four sequential “bursts” of equal-rate, sinusoidal AM (i.e., a coherent sequence of AM) applied to either a high-frequency pure-tone or broadband-noise carrier. On each single-interval detection trial, a multiburst, multicomponent masker modulator was applied to the same carrier as the target. The target- and/or masker-AM components associated with each burst of AM were gated on and off simultaneously; the carrier, however, was continuous (within each observation interval). In the MBS condition, the masker was created by randomly selecting two AM rates from a wide range prior to each stimulus presentation and then using the same two rates to generate a two-component masker modulator for each burst of the sequence. In the MBD condition, by contrast, two random rates were selected for each burst independently. Thus, the two AM components associated with the masker were either coherent (MBS condition) or incoherent (MBD condition) across bursts. Following the same logic outlined by Kidd et al., we reasoned that, if relative-coherence-based grouping and segregation mechanisms operate in the AM domain, the detectability of the target would improve in the MBD condition relative to the MBS condition because, in the MBD condition, the target, when present, formed the only coherent sequence of AM. As such, it would be perceived as segregated from the incoherent masker. In the MBS condition, by contrast, all AM components (target and/or masker) formed coherent sequences of AM, and, as such, would be grouped together perceptually, making selective attention to the target much more difficult than under MBD conditions.

4.3. Experiment 1

The purpose of Experiment 1 was to test the basic hypothesis that the relative coherence of competing AM components can be an important factor influencing the amount of modulation IM in a multisource mixture. Target-AM-detection thresholds were measured under MBS and MBD conditions (adapted to the AM domain) with the expectation that thresholds would be lower (better) in the MBD condition than in the MBS condition for the reasons outlined above.

4.3.1. Methods

4.3.1.1. Listeners

Eight listeners, including the author (C.C.), participated in Experiment 1 (five females; 21-32 years; mean = 23 years). All listeners had pure-tone air-conduction thresholds at or below 25 dB HL at octave frequencies from 250 to 8000 Hz in both ears. All listeners, save for the author, were uninformed regarding the purpose of the experiment, and received compensation for their participation. Recruitment and use of human subjects protocols were approved by the Boston University Charles River Campus Institutional Review Board.

4.3.1.2. Stimuli and procedures

A single-interval, two-alternative forced-choice (yes-no) procedure was used to measure the detectability of the target under MBS and MBD conditions. In both conditions, the target was a four-burst sequence of 32-Hz sinusoidal AM (SAM) applied to either a high-frequency (7500-Hz) pure-tone or broadband-noise carrier. In both conditions, the masker was a four-burst sequence of two-component AM applied simultaneously to the same carrier as the target. In the MBS condition, the two masker-AM components were the same on each burst of each trial but were selected at random on each trial. This across-trial randomization procedure was intended to produce large amounts of masker AM-rate uncertainty and thus large amounts of modulation IM (Conroy and Kidd, 2021). In the MBD condition, the two masker-AM components were selected at random on each burst (i.e., varied randomly across both bursts and trials). Thus, in the MBS condition, the two masker-AM components were coherent across bursts, whereas, in the MBD condition, they were incoherent across bursts, leaving the target, when present, as the only coherent sequence of AM. The selection of the two masker-AM components on each burst (MBD condition) or trial (MBS condition) was accomplished using the following procedure: first, two "base"-masker rates were selected at random (without replacement) from a limited set of 21-possible rates, log-spaced between 4 and 128 Hz. Rates falling within a protected zone of roughly +/-25% surrounding the AM rate of the target (32 Hz) were excluded. Next, two "actual"-masker rates were selected at random from two uniform distributions spanning a factor of +/-1.1 around each base-rate value. We tested both pure-tone and broadband-noise carriers under the assumption that the broadband-noise carrier would produce more within-AM-channel masking of the target- and/or masker-AM components due its intrinsic amplitude fluctuations (e.g., Dau et al., 1999). Because such within-AM-channel masking could, in turn, affect the internal representation of coherence, we reasoned that the intrinsic fluctuations of the noise carrier could obscure coherence sensitivity. Hence, we tested the pure-tone carrier so as to eliminate the influence of such masking. The choice of 7500 Hz for the pure-tone carrier was motivated by a desire to mitigate the influence of sideband detection/perception as a cue in our task (e.g., Kohlrausch et al., 2000), in particular given the dynamically changing nature of the sidebands of our stimuli (i.e., the sidebands changed from burst-to-burst on each trial).

The total duration of the carrier was 2000 ms. Each burst was 500 ms. Thus, the bursts were contiguous and, taken together, occupied the full duration of the carrier. In both the MBS and MBD conditions, the phases of all AM components (target and/or masker) were separately selected at random on each burst from a uniform distribution spanning 0 to 2π. The masker- and/or target-AM components associated with each burst were gated on and off simultaneously using 5-ms cosine-squared ramps, combined additively, and raised before multiplication with the 2000-ms carrier. Following the application of modulation, all stimuli were ramped using 5-ms cosine-squared ramps and scaled to an overall level of 55-dB SPL. Prior to ramping/scaling (but following the application of modulation), the broadband-noise-carrier stimuli were bandpass filtered between 80 and 8000 Hz using an eighth-order Butterworth bandpass filter.

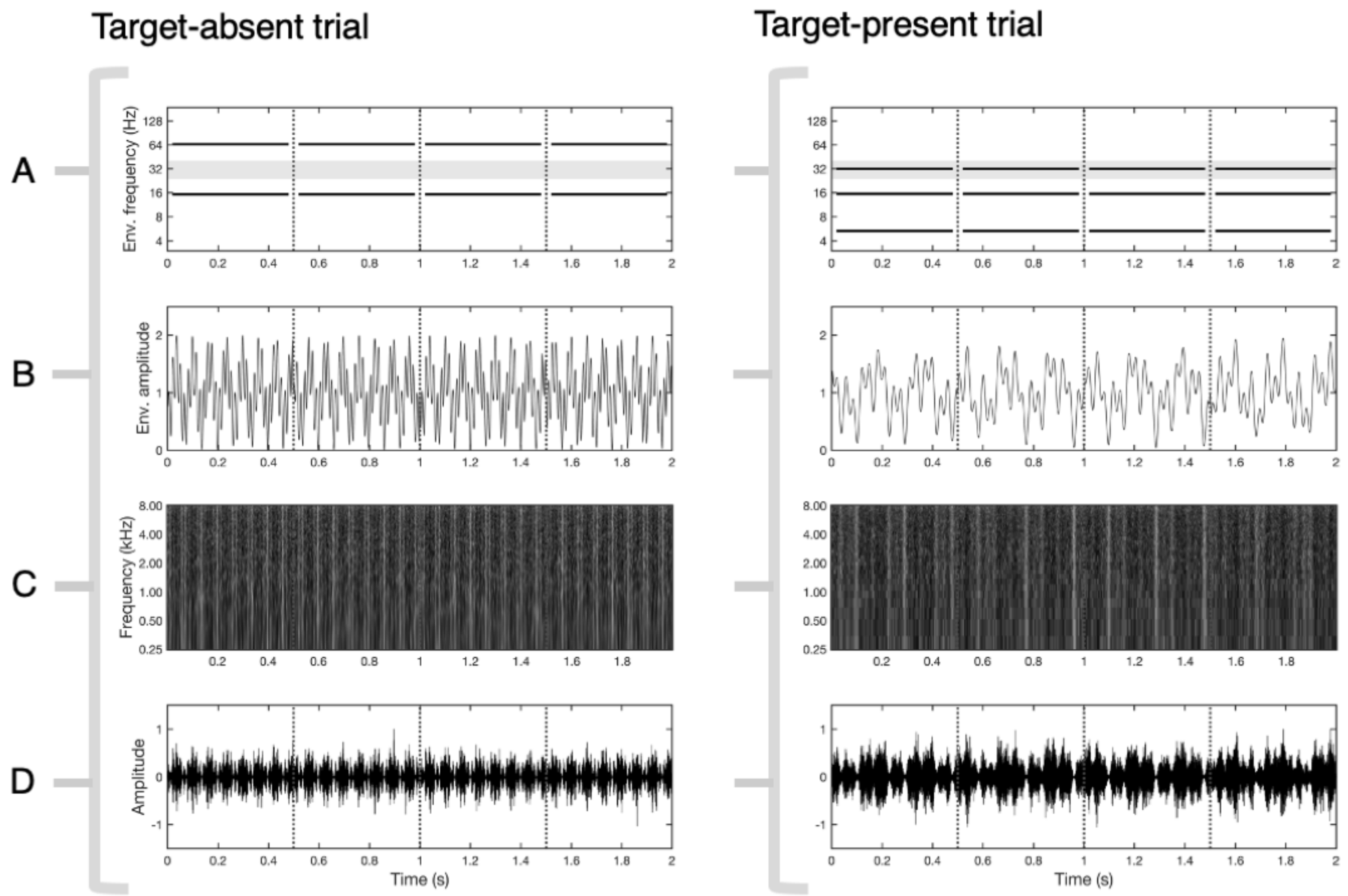

Figure 4.1: Different views of the stimuli for the multiple-bursts-same (MBS) condition. The left column shows a target-absent trial, whereas the right column shows a target-present trial. The different rows within each column show different aspects of the stimuli for the same target-absent or target-present trial. **A:** Idealized spectrogram representations of the envelopes. Dotted lines demarcate bursts, while the shaded gray regions demarcate the protected zone. **B:** Time-domain representations of the envelopes shown in row A. **C:** Spectrogram representations of the stimuli that were obtained when the envelopes shown in rows A/B were applied to a broadband-noise carrier. Darker values indicate greater stimulus energy. **D:** Time-domain representations of the stimuli shown in the spectrograms in row C. Note that, in this figure, $m_t = 0.33$ (and thus m_m does as well).

Figure 4.1 shows different views of the stimuli for the MBS condition. Each row of the left column shows a different view of a target-absent stimulus, whereas each row of the right column shows a different view of a target-present stimulus. Row A shows idealized spectrogram representations of the envelopes for the two stimulus types (i.e., the envelopes prior to multiplication with the carrier). The different AM components (target and/or masker) associated with each burst of AM are shown as solid lines, bursts are demarcated by dotted lines, and the protected zone is shown in gray. Note that the gaps between the components associated with consecutive bursts of AM are meant to illustrate the ramping of the envelope; in the experiment, the bursts were contiguous. Rows B-D show additional views of the stimuli for the MBS condition, constructed using the envelopes shown in row A. Of particular note is row C, which shows the spectrogram representations of the stimuli that were obtained when the envelopes shown in row A were applied to a broadband-noise carrier. Note how, despite the conceptual parallels with the frequency-domain version of the MBS-MBD experiment, the stimuli used in the current study were quite different qualitatively. Figure 4.2 shows different views of the stimuli for the MBD condition, plotted according to the same conventions as the MBS stimuli shown in Figure 4.1. Note that the crucial difference between the MBS and MBD stimuli was that, in the MBD condition, instead of fixing the two masker-AM components across all bursts of each sequence (Figure 4.1, row A), the two masker-AM components were selected at random on each burst (Figure 4.2, row A).

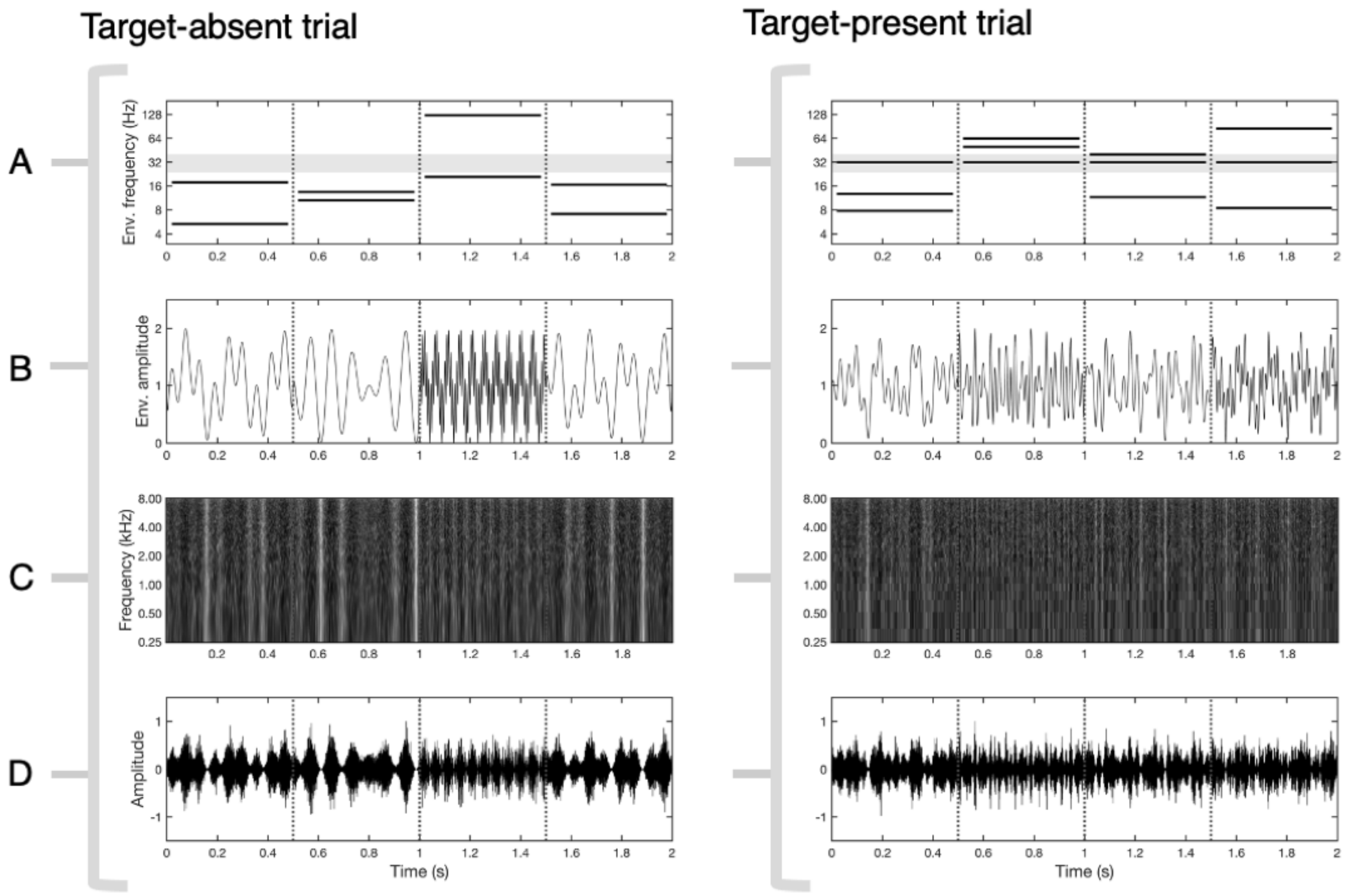

Figure 4.2: As in Figure 4.1 but for the multiple-bursts-different (MBD) stimuli.

A two-down, one-up adaptive-tracking algorithm was used to adjust the modulation depth of the target to find the minimum (relative; see below) modulation depth that was required to support threshold-level detection performance, i.e., 71% correct responses in the single-interval task (Levitt, 1971). Correct answer feedback (correct/incorrect) was provided on all trials. The modulation indices of the two masker-AM components were always equal on a given trial but differed across trials. More specifically, the sum of the modulation indices of the target, m_t, and each masker component, m_m, equaled one on all trials; the adaptive-tracking algorithm adjusted m_t; thus, as m_t decreased, m_m increased (cf. Buss et al., 2019). We adjusted the relative depths of the target and masker in this manner because the task was quite difficult and we found that fixing the modulation depth of the masker at a suprathreshold level and adapting on target depth, for example, was too difficult for some listeners to achieve above-chance performance even at the highest possible target depth. It also gave the listeners a chance to hear the target sequence in the presence of a low-depth masker for some number of trials at the beginning of each adaptive track (see below) and so each adaptive track served to refamiliarize the listener with the target sequence alone.

At the beginning of each adaptive track, $m_t = 1$ and $m_m = 0$; m_t was adjusted in dB (i.e., $20\log m_t$). An initial step size of 1 dB was used, which was reduced to 0.5 dB following the first four reversals. The adaptive track was terminated after 8 total reversals. If the adaptive-tracking algorithm called for a value of $m_t > 1$, m_t was set to one for that trial. The target was present with a 0.5 *a priori* probability on each trial. A masker was present on all trials. Data collection was completed in two experimental sessions (sometimes on the same day, sometimes on different days). The different carrier types were tested in different experimental sessions. Within each experimental session, MBS and MBD blocks (i.e., adaptive tracks) were tested in alternation, always beginning with an MBS block. Listeners completed 10 blocks for each carrier type and condition, the first five of which were considered practice, the final five of which were used to estimate thresholds. (Note that the author did not complete any formal practice blocks but had extensive informal experience with the stimuli and task.) Four of the eight listeners completed the noise-carrier session first. Thresholds were expressed in terms of the target-to-masker ratio (TMR) at threshold. Threshold TMRs were defined as the ratio, in dB, of the modulation index of the target to the sum of the modulation indices of the two masker components, i.e., $20\log(m_t/2m_m)$. A threshold TMR was estimated for each block by taking the mean of the TMR values across the final four reversals. The five threshold estimates provided by each listener were then averaged to obtain the final threshold estimate for each carrier type and condition

Stimuli were generated digitally at a sampling rate of 44100 Hz using custom scripts (MATLAB, MathWorks, Natick, MA), routed through a 24-bit sound card (RME Digiface, Haimhausen, Germany), and presented monaurally (left ear) through a pair of circumaural earphones (Sennheiser, HD280 Pro, Wedemark, Germany). Listeners were tested individually while seated in a double-walled, sound-attenuated booth (Industrial Acoustics Company, North Aurora, IL).

4.3.2. Results and discussion

Figure 4.3 shows the results of Experiment 1. In the main panel, group-mean threshold TMRs are plotted as a function of condition (MBS versus MBD) for both the high-frequency pure-tone (white symbols) and broadband-noise (black symbols) carriers. The inset panel shows the group-mean MBD benefit (MBS−MBD threshold TMR, in dB). In both panels, error bars show standard errors of the mean (SEM).

On average, thresholds were lower (better) for the noise carrier than the tone carrier and for the MBD condition than the MBS condition, suggesting that, for both carrier types, the relative coherence of the target- and masker-AM components influenced the detectability of the target. A two-way repeated-measures analysis of variance confirmed that the main effects of carrier type [$F(1,7) = 6.507$, $p = .038$] and condition [$F(1,7) = 9.319$, $p = .019$] were significant, but that their interaction was not [$F(1,7) = 0.072$, $p = .796$]. A single-sample *t*-test (single-sided; performed on the MBD benefits after averaging across carrier types) confirmed that the MBD benefit was significantly greater than zero [$t(7) = 3.053$, $p =.009$]. Thus, the results of Experiment 1 were consistent with the hypothesis that the relative coherence of competing AM components can be an important factor influencing modulation IM: in the MBD condition, the overall amount of masking experienced by the listeners, and thus, presumably, the amount of modulation IM, was reduced relative to the MBS condition, potentially reflecting the perceptual segregation of the target based on its relative coherence with the masker. Such perceptual segregation, in turn, could have facilitated selective attention to the target, suggesting more generally that perceptual grouping and segregation based on relative coherence may be one mechanism by which listeners are able to selectively attend to, and extract information from, a particular region of a stimulus' AM spectrum under conditions of high masker AM-rate uncertainty (modulation IM).

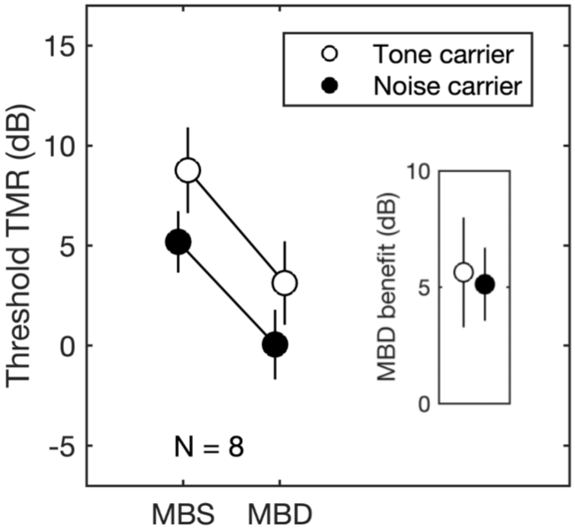

Figure 4.3: Main: Group-mean threshold target-to-masker ratios (TMRs) for each carrier type and condition. White symbols show the TMRs for the tone carrier, whereas black symbols show the TMRs for the noise carrier. Lower threshold TMR values indicate better performance. **Inset:** Group-mean MBD benefit (MBS−MBD threshold TMR, in dB) for the two carrier types. Error bars in both panels are standard errors of the mean (SEM).

The notion that perceptual segregation of the target and/or coherence-sensitive mechanisms contributed to the MBD benefit evident in Figure 4.3, however, warrants considerable caution at this juncture, as other interpretations of the MBD benefit remain equally viable. In both the MBS and MBD conditions, for example, threshold TMRs were greater, on average, than 0 dB (not to mention, -6 dB), suggesting that, for some listeners, the target dominated the "mixture" (i.e., overall envelope shape) at detection threshold. This suggests, in turn, that segregation of the target based on relative coherence, or indeed, selective attention to the target at all, may not have been necessary to support detection. One way to view this interpretation is in terms of the “envelope of the envelope,” or the “venelope,” of the stimuli (cf. Lorenzi et al., 2001; Ewert et al., 2002; Füllgrabe et al., 2005). The addition of the target tended to reduce venelope fluctuations, in particular in the MBD condition (cf. Figures 4.1 and 4.2), and this could have been used as a detection cue. An alternative view of the same interpretation is one in terms of texture perception (e.g., McDermott and Simoncelli, 2011; McDermott et al., 2013; McWalter and Dau, 2017): the addition of the target tended to produce a “smoother” auditory texture, and this perceptual difference (i.e., of the perceived smoothness of the envelope between masker-alone and target-plus-masker trials), again, could have been used a detection cue.

Alternatively, it is possible that, in both the MBS and MBD conditions, detection of the target was based on the single “best burst” of each sequence, i.e., the burst on which the internal representation of the target was least obscured by the masker. One way to view this interpretation is in terms of AM channels, i.e., psychophysical channels tuned to different AM rates (cf. Dau et al., 1997; Ewert and Dau, 2000). Within the framework of AM channels, a “best-burst interpretation” posits that differences in the amount of within-AM-channel (i.e., modulation) masking produced by the different, randomly selected masker-component pairs was responsible for the MBD re MBS benefit. Recall that, in the MBS condition, only one randomly selected masker-component pair was tested across the four bursts of each sequence, whereas, in the MBD condition, four, randomly selected masker-component pairs were tested on each sequence. If at least one of these masker-component pairs produced relatively little within-AM-channel masking, the target could have “popped out” on that burst, supporting detection. By this view, then, an MBD benefit is predicted simply on the basis of the statistics of the stimuli. Previous estimates of AM-channel bandwidths obtained under low-uncertainty conditions suggest that the tuning of the putative AM channels may be rather broad; indeed, much broader than is implied by the width of the protected zone employed here (e.g., Dau et al., 1997; Ewert and Dau, 2000; Wojtczak and Viemeister, 2005; Moore et al., 2009; Wojtczak, 2011). As such, it is certainly possible that at least some of the maskers tested in the current experiment produced at least some within-AM-channel masking, and that such masking was reduced in the MBD condition. Note, moreover, that detection of the target on the basis of the best burst of each sequence also could reflect differences across maskers in terms of the extent to which they produced IM, and so an MBD re MBS benefit based on a best-burst detection strategy also could reflect a reduction of IM. We return to a best-burst interpretation below.

Neither a venelope-based nor best-burst-based interpretation of the MBD benefit can account for the finding that thresholds were lower for the noise carrier than for the tone carrier for both the MBS and MBD conditions, an unexpected finding that runs counter to the prediction that the intrinsic fluctuations of the noise carrier should have reduced the detectability of the target (relative to the tone carrier condition) via within-AM-channel masking (cf. Dau et al., 1999). A relative-coherence-based interpretation of the MBD benefit, however, may help to explain this unexpected finding. While the intrinsic fluctuations of the noise carrier may have produced some within-AM-channel masking, they also introduced incoherent AM to different regions of the AM spectrum (considering the AM spectrum on a burst-by-burst basis). In principle, this incoherent background AM could have been used to support the perceptual segregation of the target, improving its detectability in turn. One way to test this idea would be to measure *unmasked* target-AM-detection thresholds under two different conditions: a frozen-noise-carrier condition, in which the same noise-carrier segment is used across all bursts of each sequence, and a random-noise-carrier condition, in which a different noise-carrier segment is used on each burst of each sequence. If performance improves for the random-carrier re frozen-carrier condition, one could argue that the relative coherence of the *imposed* target and the *intrinsic* carrier fluctuations was a factor influencing the detectability of the target. Informal pilot data taken from the author (C.C.) suggest that this relative-coherence-based prediction may bear out. The same best-burst logic outlined above, however, would apply under these conditions as well (i.e., randomizing the carrier across bursts would increase the probability of at least one burst producing relatively little masking of the target) and so it is important to determine the extent to which a best-burst interpretation can account for the MBD benefit shown in Figure 4.3. Experiment 2, therefore, examined the burst-burst interpretation in more detail.

4.4. Experiment 2

A best-burst interpretation of the MBD benefit assumes that, under multiburst conditions (i.e., either MBS or MBD conditions), the bursts are independent, meaning that the detectability of the target is unaffected by across-burst interactions. This is in contrast to a relative-coherence-based interpretation, to which across-burst interactions are essential (i.e., coherence sensitivity is defined as sensitivity to the relatively rapid and repeated presentation of a particular AM component). Accordingly, the best-burst interpretation, but not a coherence-based interpretation, suggests that the detectability of the target under multiburst conditions should reflect the detectability of the target on the best burst of each sequence—independent of all others. This implies that if the detectability of the target is known for each burst of each sequence, performance can be predicted based on performance for the "best burst" alone. Such prediction was the goal of Experiment 2. With this goal in mind, the MBD condition was tested again (with important differences in masker generation, noted below), as was an SB random-masker condition. In the SB random-masker condition, a limited set of 10 masker-component pairs was used to generate the masker modulator on each trial. On each trial of the MBD condition, the multiburst, multicomponent masker modulator also was generated using the same limited set of 10 masker-component pairs, with one of the 10 pairs randomly selected to generate the two-component masker modulator for each burst. In the SB condition, performance was measured for each pair separately and then used to predict performance in the MBD condition assuming a best-burst detection strategy. We hypothesized that if a best-burst detection strategy was responsible for producing the MBD benefit in Experiment 1, the best-burst predictions would capture the magnitude of the effect in Experiment 2. Furthermore, we obtained estimates of within-AM-channel masking for the 10 masker-component pairs and compared these predictions to the psychophysical results. We hypothesized that if the detectability of the target in both the

SB and MBD conditions was related to within-AM-channel masking, the pattern of performance across masker-component pairs in the SB condition would be similar to the predicted pattern of within-AM-channel masking.

4.4.1. Methods

4.4.1.1. Listeners

Five listeners participated in Experiment 2 (four females; 18-25 years; mean = 22 years), one of whom had participated in Experiment 1. The author did not participate in Experiment 2. All listeners had pure-tone air-conduction thresholds at or below 20 dB HL at octave frequencies from 250 to 8000 Hz in both ears, were uninformed regarding the purpose of the experiment, and received compensation for their participation. Recruitment and use of human subjects protocols were approved by the Boston University Charles River Campus Institutional Review Board.

4.4.1.2. Stimuli and procedures

The details of the stimuli and procedures were very similar to Experiment 1, with some notable exceptions. A single-interval, two-alternative forced-choice procedure was used to measure the detectability of the target under SB and MBD conditions. In the SB condition, the target was a single, 500-ms burst of 32-Hz SAM, whereas, in the MBD condition, it was the same as in Experiment 1 (i.e., four, temporally contiguous, 500-ms bursts of 32-Hz SAM). Only the broadband-noise carrier was tested. In the SB condition, the carrier was 500 ms, whereas, in the MBD condition, it was 2000 ms. Thus, in both conditions, the target, when present, occupied the full duration of the carrier.

On each trial of the SB condition and each burst of the MBD condition, a 500-ms, two-component masker modulator was applied to the same carrier as the target. By contrast with Experiment 1, however, a limited set of 10 masker-component pairs was used to generate the masker modulators in both conditions. Idealized envelope-spectrum representations of the 10 masker-component pairs are shown in Figure 4.4. The 10 pairs were generated once prior to the start of the experiment using the same procedure as was used to select the two base-masker components for each burst in the MBD condition in Experiment 1 (see Section 4.3.1.2). The same set of 10 masker-component pairs were used for all listeners. On each trial of the SB condition, one pair was randomly selected from the set of 10 on each trial and then each component was jittered using the same procedure as was used to select the two actual-masker components in Experiment 1 (again, see Section 4.3.1.2). On each burst of the MBD condition, one pair was randomly selected (with replacement) from the set of 10 and then jittered in a similar manner. As in Experiment 1, the phases of all AM components (target and/or masker) were separately selected at random on burst and all AM components associated with each burst were summed, gated on and off simultaneously using 5-ms cosine-squared ramps, and raised before multiplication with the broadband-noise carrier. Following the application of modulation, all stimuli were filtered between 80 and 8000 Hz using an eighth-order Butterworth bandpass filter, gated on and off using 5-ms cosine-squared ramps, and scaled to an overall level of 55-dB SPL. The target was present with a 0.5 *a priori* probability on each trial. A masker was present on all trials.

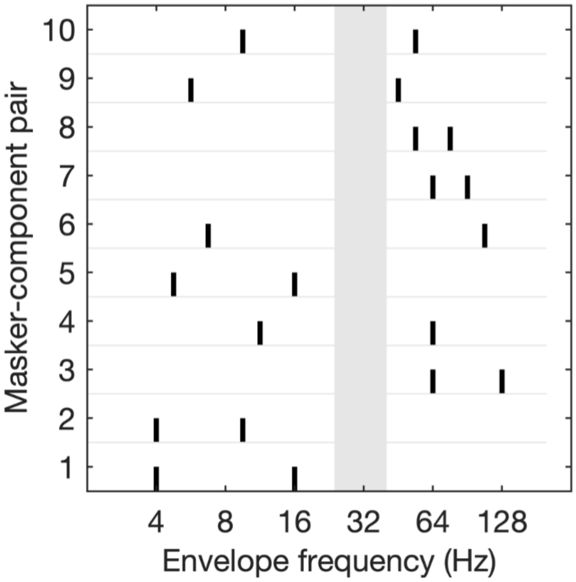

Figure 4.4: Idealized envelope-spectrum representation of the 10 masker-component pairs that were used to construct the maskers for the SB and MBD conditions. In the SB condition, one pair was randomly selected on each trial to construct a two-component masker modulator. In the MBD condition, one pair was randomly selected on each burst. The shaded gray region illustrates the protected zone centered on the AM rate of the target, i.e., 32 Hz.

The proportion of correct responses registered by each listener in the single-interval detection task was measured in 30-trial blocks. Correct answer feedback (correct/incorrect) was provided on all trials. SB and MBD trials were randomly intermixed within each block, with 15 trials per condition tested in each block. Unlike in Experiment 1, where an adaptive-tracking procedure was used to adjust the relative depths of the target and masker AM to find a threshold TMR for each listener and condition, in Experiment 2, the detectability of the target was assessed at a fixed relative depth. As in Experiment 1, the modulation indices of all AM components on a given trial always summed to one; m_t was always 0.33. Thus, on target-present trials, $m_m = 0.33$, whereas, on target-absent trials, $m_m = 0.50$. Prior to beginning data collection, listeners completed two blocks of unmasked target-modulation detection to become familiar with the target. During the unmasked target-detection blocks, both SB and MBD targets were tested using stimuli that were identical to the main experiment except that $m_m = 0.0$ (note, however, that m_t was fixed at 0.33). Following this initial familiarization period, the listeners completed either 34 (one listener) or 60 (four listeners) 30-trial blocks (depending on subject availability), the first half of which were considered practice (as in Experiment 1). Overall proportion correct scores were calculated for both conditions as well as for each of the 10 masker-component pairs in the SB condition.

4..4.1.3. Best-burst predictions

For each listener, predicted performance assuming a best-burst detection strategy was calculated in the following manner: first, all SB trials on which a particular masker-component pair was tested were pooled, and a proportion correct score was calculated. Next, the exact sequence of trials (SB and MBD conditions) completed by each listener was simulated 1000 times, with the probability of success on each trial determined by the "best burst" on that trial, i.e., the burst on which the masker-component pair that was tested yielded the highest proportion-correct score in the SB condition. Thus, for SB trials, the probability of success on a given trial was equal to the proportion correct for the masker-component pair that was tested on that trial. For MBD trials, by contrast, the probability of success was determined by one of the up to four-possible proportion correct scores associated with the up to four masker-component pairs tested on that trial. The simulated proportion correct over all 1000 simulations of the experiment was then taken as the predicted proportion-correct score for each listener for each condition.

4.4.1.4. Within-channel masking predictions

In Sec. II.B we introduced the notion that the best burst of each sequence may have been the one on which the masker modulator produced the least within-AM-channel masking. This suggests that, if a best-burst detection strategy was responsible for producing the MBD benefit, the pattern of proportion-correct scores across masker-component pairs in the SB condition should have been consistent with the pattern of within-AM-channel masking produced by the different masker-component pairs. To examine this prediction in more detail, we obtained estimates of the modulation power at the output of the target-AM channel (i.e., an AM channel tuned to 32 Hz) for each of the 10 masker-component pairs using a simplified version of the envelope power spectrum model (EPSM) of Ewert and Dau (2000). The logic was simple: larger modulation-power estimates should correspond to masker-component pairs that more strongly stimulated the target channel and thus produced the most within-AM-channel masking of the target. To obtain each modulation-power estimate, we created a 10-second-long sample of an SB masker for each masker-component pair on a target-absent trial (i.e., $m_m = 0.5$). The magnitude spectrum of the masker (i.e., the ac-coupled envelope) was then multiplied by the magnitude spectrum of a second-order Butterworth bandpass filter with a center frequency of 32 Hz and a quality factor (Q; center frequency divided by bandwidth) of 1 (i.e., the putative target-AM channel). A modulation-power estimate was then obtained by integrating the filtered masker-AM spectrum across rate (cf. Ewert and Dau, 2000).

4.4.2. Results and discussion

The results of Experiment 2 are shown in Figure 4.5. In the main panel, group-mean proportion-correct scores are plotted as a function of condition (SB versus MBD; white symbols) along with the group-mean predictions based on a best-burst detection strategy (black symbols). The inset panel shows both the observed and predicted group-mean MBD benefit (proportion correct in the MBD condition−proportion correct in the SB condition). In both panels, error bars show SEM.

On average, performance improved in the MBD condition relative to the SB condition, yielding a small, yet significantly positive, MBD benefit [single-sample *t*-test (single-sided), $t(4) = 3.626$, $p = .011$]. This is consistent with the suggestion that the MBD manipulation tapped coherence-sensitive mechanisms that effected a perceptual segregation of the target from the masker and, potentially, facilitated selective attention to the target and/or suppression of the masker. The predicted MBD benefit assuming a best-burst detection strategy, however, easily captured the magnitude of the effect, suggesting that the MBD benefit observed psychophysically may *not* have been related to the relative coherence of the target- and masker-AM components. Indeed, an independent-samples t-test (two-sided) indicated that the predicted MBD benefit was significantly greater than the MBD benefit observed psychophysically [$t(4) = 5.006$, $p = .008$]. The results of Experiment 2, therefore, were consistent with the hypothesis that the MBD benefit in Experiment 1 (and here) reflected the improved detectability of the target associated with the statistics of the stimuli (i.e., a best-burst detection strategy), rather than a perceptual effect related to coherence sensitivity *per se*.

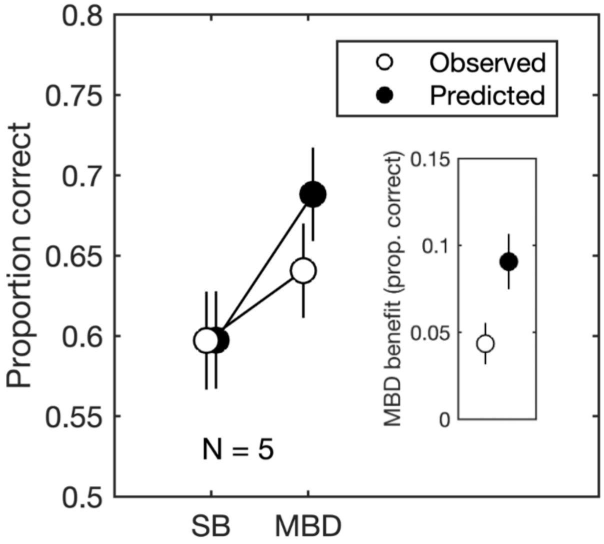

Figure 4.5: Main: Group-mean proportion-correct scores for both the single-burst (SB) and MBD conditions (white symbols) as well as group-mean best-burst predictions for the same conditions (black symbols). **Inset:** Group-mean MBD benefit (proportion correct in the MBD condition−proportion correct in the SB condition; white symbol) as well as the predicted MBD benefit assuming a best-burst strategy. Error bars in both panels are SEM.

A comparison of the SB data to the within-AM-channel-masking predictions (Figure 4.6), however, suggests that the single-burst detection strategy, if used, did not yield an MBD benefit via a reduction of within-AM-channel masking and, instead, may have yielded a reduction of modulation IM. Figure 4.6 shows the group-mean proportion-correct scores for the different masker-component pairs in the SB condition plotted as a function of the modulation-power at the output of the putative target-AM channel on masker-alone trials for each pair. Different symbols show the results for different masker-component pairs (note that the component-pair numbers given in the legend of Figure 4.6 correspond to the component-pair numbers given in Figure 4.4). Error bars are SEM. The solid line in Figure 4.6 is a least-squares linear-regression line fit to the full data set, i.e., the individual proportion-correct scores plotted as a function of the masker-modulator power estimates. As can be seen from Figure 4.6, while there was a *slight* trends towards poorer performance for masker modulators that more strongly stimulated the target channel (i.e., a slightly negative-sloping regression line), the correlation between the observed proportion-correct scores and masker-modulator power estimates was weak (note the large error bars) and failed to reach statistical significance at the $p < .05$ level ($r = .206$, $p = .152$). This suggests that differences in the amount of within-AM-channel masking across the different masker modulators was not the primary factor influencing performance in the SB condition. This suggests, in turn, that other forms of masking (i.e., modulation IM) contributed to the differences across maskers and that factors other than a reduction of within-AM-channel masking, such as a reduction of modulation IM, contributed to the MBD benefit in Experiment 2 (and perhaps in Experiment 1 as well).

Finally, it is worth noting that while the predictions shown in Figure 4.5 suggest that a best-burst detection strategy *could* produce an MBD benefit, it is not necessarily the case that such a strategy was used by the listeners and so, at this point, coherence sensitivity cannot be ruled out as a contributor to the MBD benefit.

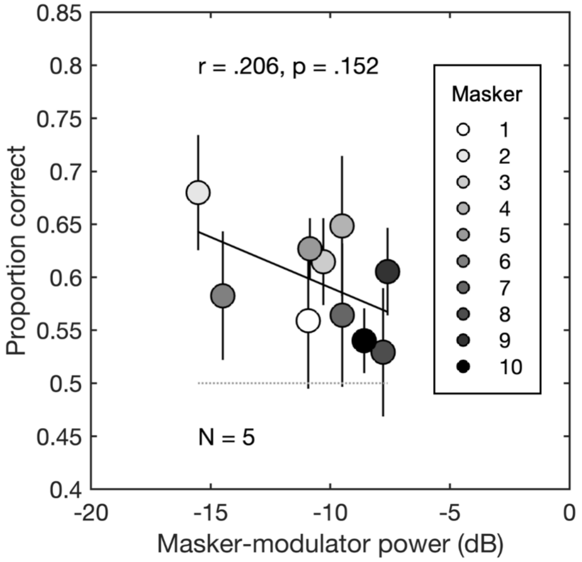

Figure 4.6: Group-mean proportion correct scores in the SB condition for each masker-component pair plotted as a function of the masker-modulator power at the output of the target AM channel. Different shades of gray correspond to different masker-component pairs, with the number in the legend corresponding to the number in Figure 4.5. Error bars are SEM. The solid line shows the best-fitting regression line to the full data set (i.e., to the individual data), with the inset correlation coefficient and *p* value also referring to the full data set. While there was a slight trend towards poorer performance for maskers that produced greater amounts of within-AM-channel masking, masker-modulator power at the output of the target-AM channel was, in general, a poor predictor of performance in the SB condition, suggesting that factors other than within-AM-channel masking contributed to the MBD benefit.

4.5. General discussion

The results reported in this chapter suggest that a target comprising a coherent sequence of AM is more easily detected when embedded in a masker comprising incoherent (MBD condition) than coherent (MBS and SB conditions) sequences of AM, even in the absence of all other cues to detection (e.g., spectral or spatial differences between the target and masker). In all conditions, the masker-AM sequences varied randomly across trials, producing large amounts of masker AM-rate uncertainty and thus, presumably, large amounts of modulation IM (Conroy and Kidd, 2021). The improvement in performance under MBD re MBS and SB conditions, then, can be viewed as a release from modulation IM, potentially reflecting the operation of perceptual grouping and segregation mechanisms sensitive to AM-domain coherence. Some aspects of the results, however, suggested that simpler mechanisms, and, in particular, a release from within-AM-channel masking (i.e., EM in the AM domain), may have contributed to the MBD benefit, suggesting that perceptual segregation of the target and/or improved selective attention to it may not have been necessary. Nevertheless, the results, taken together, are consistent with the suggestion that perceptual grouping and segregation based on relative coherence may be one way that the auditory system copes with IM arising from listener uncertainty along different acoustic dimensions. Previous studies using MBS-MBD-type manipulations have suggested as much for the frequency domain; the current study suggests as much for the AM domain.

Yet while the results of the current study highlight important similarities between different types of IM, they also highlight important differences. A particularly notable difference is the size, in a *d'* sense, of the MBD benefits reported here relative to the MBD benefits that have been reported previously in the frequency domain. More specifically, the MBD benefits reported here were rather meager, whereas frequency-domain MBD benefits can be enormous (e.g., Kidd et al., 1994; Oxenham et al., 2003; Kidd et al., 2003; Durlach et al., 2003b). There are at least three potential explanations for the difference. First, it is possible that, in the current study, the MBD manipulation did not yield a perceptual segregation of the target or facilitate selective attention to it and instead was attributable to one of the other factors discussed above (e.g., detection based on overall envelope shape or a reduction of EM in the AM domain). For MBS-MBD-type experiments conducted in the frequency domain, there are empirical data that strongly suggest that a reduction of within-frequency-channel masking (i.e., EM) does not contribute appreciably to the MBD benefits that are observed (e.g., Kidd et al., 2003), although increasing the width of the protected zone in such experiments does tend to improve performance overall (e.g., Micheyl et al., 2007). Regardless, in the frequency domain, hit responses in MBD-type conditions typically are characterized by a robust perceptual “pop out” of the target: as Kidd et al. (1995) put it, a coherent sequence of pure tones embedded in incoherent masker sequences often is perceived as “‘sticking’ together to form a pattern” (pp. 3782), which is clearly segregated from the randomly varying background. Our (author C.C.) own impressions suggest that, in the current study, no such perceptual pop-out occurred, or if it did, it was relatively weak, and that the “sticking-togetherness” of the target (i.e., the subjective impression of the target as a single auditory object) was much weaker than in the frequency domain. Obtaining subjective estimates of perceptual segregation in future studies using similar stimuli could help to shed light on this issue.

A second reason why the MBD benefits observed in the current study may have been relatively small with respect to previous frequency-domain MBS-MBD-type experiments is that the random-rate masker modulators may have been more perceptually similar *to each other* than the random-flanker complexes typically employed in MBS-MBD type experiments. If this was the case, the randomization of the masker across trials in the SB and MBS conditions would have produced relatively little listener uncertainty/modulation IM and thus there would have been relatively little modulation

IM to reduce via relative-coherence-based perceptual segregation/improved selective attention. A similar interpretation was offered by Fan et al. (2008) in attempting to account for differences in the amount of IM produced by masker randomization along the spatial dimension in an SB random-masker-type experiment. Indeed, AM-rate discrimination and identification data obtained under low-uncertainty conditions (e.g., Hanna, 1992) suggest that the perceptual space occupied by the maskers employed in the current study was relatively constricted, and thus the notion that randomization of the maskers produced relatively little IM seems plausible. On the other hand, perceptual similarity of the maskers with each other suggests their similarity *with the target*, and such similarity (i.e., between a target and masker) can greatly *exacerbate* the effects of IM (Neff, 1995; Kidd et al., 2002; Durlach et al., 2003b; Lee and Richards, 2011). It is therefore possible that, even if the effects of uncertainty were small, the effects of similarity were such that the overall amount of modulation IM remained high.

A third reason why the MBD benefits reported here may have been relatively small compared to those observed under frequency IM is that the MBD manipulation in the current study provided only a partial release from modulation IM. Indeed, the individual listeners were highly variable in their overall performance (again, note the size of the error bars in Figure 4.6), including in the “low-IM” MBD condition. If performance in this condition was relatively uncontaminated by modulation IM, we would expect relatively little individual variability in this condition (cf. Oxenham et al., 2003). Indeed, there are myriad factors, in particular at the level of the stimuli, that also could account for a lack of a full release from IM (e.g., burst duration, total number of bursts, the number of masker components in each burst, etc.). Thus, further work exploring the stimulus space parametrically could be of interest with respect to determining the full range of stimulus factors that can produce/reduce modulation IM.

4.6. Summary and conclusions

The results of the current study suggest that the relative coherence of competing AM components may be an important factor influencing their perceptual grouping and segregation, in particular under conditions high masker AM-rate uncertainty (modulation IM). This is consistent with the notion that grouping and segregation based on relative coherence is a general perceptual principle that applies across feature dimensions and thus with the notion that perceptual grouping and segregation based on relative coherence can be an important factor influencing different types of IM (i.e., frequency IM and modulation IM). The results of the current study, however, also highlighted important differences between different types of IM, as the magnitude of the putative perceptual effects were rather small, and could be explained by nonperceptual (e.g., within-AM-channel masking) or alternative perceptual (e.g., texture perception) factors. Additional work, therefore, is required to better understand the effects reported here.

5. GENERAL DISCUSSION

The broad goal of the work reported in this dissertation was to better understand how uncertainty regarding the amplitude-modulation (AM) content of an acoustic mixture affects the ability of human listeners to selectively attend to, and extract information from, particular regions of a stimulus' AM spectrum, in particular under conditions of ambiguous, distracting, or confusing AM input. The key findings can be summarized as follows: In Chapter 2, we found that uncertainty regarding the AM spectrum of a masker can have an adverse effect on the detectability of target AM, consistent with the notion that such uncertainty made it difficult for some listeners to selectively attend to, and extract information from, the regions of the AM spectrum associated with the target. This, we argued, was consistent with the suggestion that the effects of uncertainty revealed by previous studies of informational masking (IM) in the frequency domain, or "frequency IM" (cf. Durlach et al., 2003), apply in the AM domain as well. In other words, we took this basic finding as evidence of "modulation IM." In Chapter 3, we found that, under similar, uncertain-masker conditions, *a priori* knowledge of the AM spectrum of the masker derived from relatively brief contextual cues could be used by listeners to obtain a release from modulation IM. Building on the results of Chapter 2, we speculated that such *a priori* knowledge of the masker may have made it easier for some listeners to focus and/or maintain selective attention on the regions of the AM spectrum associated with the target. In Chapter 4 we found that stimulus manipulations intended to yield a relatively "automatic" perceptual segregation of the target AM also could provide a release from modulation IM, suggesting that both bottom-up (i.e., stimulus-driven) and top-down-attentional mechanisms can influence the overall amount of modulation IM in a mixture. In all three chapters, however, caveats and, at times, equally viable alternative interpretations, required us to temper these sweeping generalizations. Nevertheless, the effects reported in all three chapters are similar, conceptually, to previous effects reported in the frequency domain, pointing up the distinct possibility that frequency IM and modulation IM may reflect the operation of similar underlying mechanisms.

As discussed in Chapter 4, however, the results also highlight important differences between IM arising from listener uncertainty along the frequency and AM-rate dimensions. Perhaps the most salient difference related to the overall *amount* of modulation IM that was measured in the different conditions tested. More specifically, the overall amounts of modulation IM that were reported in the preceding chapters were much smaller, on average, than the overall amounts of frequency IM that have been reported in previous studies, the implication here being that, in a dynamic, multisource mixture, frequency IM may have a relatively greater influence on a listeners behavior that modulation IM. Take, for example, the experiment reported in Chapter 2. In that experiment, the group-mean amount of modulation IM across the eight listeners tested (indexed as a target-AM-detection-threshold increase under high- relative to low-IM conditions) was 8 dB. For individual listeners, it ranged from 3 to 12 dB. These values are in marked contrast to the much larger values that have been reported in analogous experiments conducted in the frequency domain (i.e., experiments where the same limited set of maskers were used to measure target-tone detection performance under low- and high-IM conditions; e.g., Wright and Saberi, 1999; Durlach et al., 2005; Leibold et al., 2010). Durlach et al. (2005), for example, reported that the group-mean amount of frequency IM across five listeners (indexed as the target-tone-detection-threshold increase under high- relative to low-IM conditions) was 13 dB, and up to 30 dB for one listener. While certain details of the stimuli and analyses used in our study undoubtedly influenced the overall amount of modulation IM that was reported, it nonetheless is clear that there are important differences between frequency IM and modulation IM in terms of the relative magnitudes of the perceptual effects they reflect.

One factor, not discussed in Chapter 4, that could have limited the amount of modulation IM reported in this dissertation is the relatively small number of AM components that were used to construct the maskers. In the experiments reported in Chapters 2 and 3, for example, the maskers were comprised of a single AM component, whereas, in the experiments reported in Chapter 4, the maskers were comprised of only two AM components. Previous studies of frequency IM have shown that increasing the number of flanker tones that comprise a multitone masker from around two to 20 can dramatically increase the amount of IM that is measured under conditions analogous to the conditions tested here (e.g., Neff and Green, 1987; Neff and Callaghan, 1988; Neff et al., 1993; Oh and Lutfi, 1998). It is possible that modulation IM similarly would increase with increasing numbers of AM components in the masker. More generally, this suggests that more complex envelopes than those used in our experiments—yet more similar to those encountered in everyday life—may produce larger effects of uncertainty than were suggested by our results. Unfortunately, testing this possibility would be difficult using stimuli similar to those used here. That is, given physical constraints associated with overmodulation, the use of multicomponent masker modulators similar to those used here would require a decrease in the modulation depth of individual components with increasing component density. Given that the perceived strength of modulation (e.g., “fluctuation strength,” “roughness,” etc.) for individual AM components tends to decrease with decreasing depth (e.g., Ritsma, 1962; Terhardt, 1968; Burns and Viemeister, 1976; Burns and Viemeister, 1981; Fastl, 1983; Wojtczak and Viemeister, 2008), it is possible that, rather than increase with number of components, modulation IM would decrease due to decreased component salience. Furthermore, many more simultaneous frequency components than AM components are resolvable by the auditory system, perhaps a consequence of the relatively broad bandwidths of the putative AM channels (cf. Dau et al., 1997; Ewert and Dau, 2000; Wojtczak and Viemeister, 2005; Moore et al., 2009; Wojtczak, 2011). As such, it is unclear how many separately resolvable AM components could be included in a masker while still producing large perceptual variations in a modulation-spectral profile.

What the above speculations and others offered in Chapter 4 have in common is that they suggest that one key difference between frequency IM and modulation IM may be related to the limitations on AM-rate resolution imposed by the “peripheral” auditory system (cf. Chapter 2). That is, poorer resolution results in less uncertainty due to smaller perceptual differences across stimuli. This interpretation is supported by the model predictions reported in Chapter 2, which suggested that internal limitations on AM-rate resolution—and thus something like energetic masking in the modulation domain—may have been an appreciable factor in the conditions reported here despite the fact that, in all conditions, a protected zone was included to mitigate such effects. Thus, additional research intended to more fully examine the extent to which limitations on AM-rate resolution contributed to the effects reported in this dissertation could be useful in understanding the relative influence of frequency IM and modulation IM in multisource mixtures. Such research could include, for example, obtaining explicit measures of modulation tuning under low-uncertainty conditions for individual listeners and then comparing these measures to measures of modulation IM in the same group. The model predictions reported in Chapter 2 suggest that listeners that exhibit poor modulation tuning should be *least* susceptible to modulation IM.

As noted in Chapter 1, the experiments reported in this dissertation were designed, primarily, to elucidate the basic aspects of modulation IM, i.e., the basic stimulus and psychological factors that can produce and reduce it. Accordingly, we used simple, nonspeech stimuli and simple psychophysical detection tasks. Nevertheless, and as implied by our preceding discussion, the results reported here may have some relevance to situations outside of the psychophysical laboratory—that is, insofar as modulation IM does as well. Consider speech, for example. The AM cues inherent to human speech are crucial contributors to speech intelligibility (e.g., Van Tasell et al., 1987; Drullman, 1994a,b; Shannon et al., 1995), yet, in everyday life they are often embedded in an acoustic background that contains AM "cues" of its own. If the AM spectrum of the background is randomly and/or unpredictably changing—as might be the case, for example, if the background is competing speech—modulation IM could act to limit speech intelligibility. Thus, the results reported here suggest that modulation IM may be one factor that affects speech intelligibility in the real world. Note that while a number of previous studies have yielded findings that are consistent with the suggestion that masking effects resulting from perceptual competition between target and masker envelopes (and/or within-AM-channel masking) may play an important role in masked speech intelligibility (e.g., Dubbelboer and Houtgast, 2008; Jørgensen and Dau, 2011; Stone et al., 2012; Stone and Moore, 2014), the results reported in this dissertation are different insofar as they suggest that listener uncertainty regarding the AM spectrum of a masker *per se* also may be an important factor to consider in the context of speech perception in multisource mixtures. Similar logic regarding the influence of modulation IM in more naturalistic settings can be applied to listening situations that do not involve speech, yet nevertheless require the selective processing of certain AM cues in a mixture (e.g., perceiving one texture embedded in another; cf. McDermott and Simoncelli, 2011).

The results reported in this dissertation also may have some relevance to certain clinical questions, or, at least, future, clinically oriented research. The fundamental importance of AM cues to listeners with cochlear implants (e.g., Fu, 2002), for example, may mean that modulation IM is a particularly relevant phenomenon for this listener group. Indeed, insofar as modulation IM is a factor in multitalker listening situations, it seems likely that the relative influence of modulation IM in such settings would be greater for listeners with cochlear implants than for listeners with normal hearing or unimplanted listeners with hearing loss. One of the most striking aspects of the performance of listeners with cochlear implants in multitalker settings is the large individual differences across listeners (e.g., Litovsky et al., 2009). While individual differences have received relatively little attention in this dissertation, they nevertheless were apparent in all three studies under both low- and high-IM conditions (see, e.g., Figure 2.2), suggesting that there were large individual differences both in listeners' susceptibility to modulation IM and in their ability to overcome it. This suggests, in turn, that individual differences in modulation IM may contribute to the large individual differences in performance often exhibited by listeners with cochlear implants, in particular in multisource settings. In terms of future research on modulation IM, this seems like a particularly fruitful avenue to explore.

Finally, we note that, consistent with our finding of large individual differences in susceptibility to modulation IM, large individual differences in frequency IM are ubiquitous in studies of that topic (e.g., Neff et al. 1993; Neff and Dethlefs, 1995; Lutfi et al., 2003a; Durlach et al., 2003b; Oxenham et al., 2003). While the full range of factors and mechanisms that underlie such differences currently is unknown, one hypothesis is that they reflect individual differences in the ability to selectively attend to, and extract information from, a particular region of an otherwise highly uncertain frequency spectrum (cf. Lutfi et al., 2003a; Durlach et al., 2005). An extension of this hypothesis is that individual differences in IM—be it frequency IM, modulation IM, or "spatial IM" (cf. Fan et al., 2008)—may reflect individual differences in a more general ability to focus attentional resources at a particular point along an acoustic dimension in the presence of dynamic or unpredictable changes in a masker. Insofar as the results reported in this dissertation suggest that different forms of IM reflect similar underlying mechanisms, they suggest that this “general attentional” hypothesis is correct. Future studies explicitly designed to probe individual differences in modulation IM, and, more importantly, the *consistency* of such differences across different acoustic dimensions (e.g., frequency, space), would help to clarify the mechanisms by which listener uncertainty and expectation affect auditory perception and communication in dynamic, multisource mixtures.

BIBLIOGRAPHY

Allen, P., and Wightman, F. (1995). Effects of signal and masker uncertainty on children's detection. *Journal of Speech, Language, and Hearing Research*, *38*(2), 503–511.

Bacon, S. P., and Grantham, D. W. (1989). Modulation masking: Effects of modulation frequency, depth, and phase. *The Journal of the Acoustical Society of America*, *85*(6), 2575–2580.

Bacon, S. P., and Grantham, D. W. (1992). Fringe effects in modulation masking. *The Journal of the Acoustical Society of America*, *91*(6), 3451–3455.

Burns, E. M., and Viemeister, N. F. (1976). Nonspectral pitch. *The Journal of the Acoustical Society of America*, *60*(4), 863–869.

Burns, E. M., and Viemeister, N. F. (1981). Played-again SAM: Further observations on the pitch of amplitude-modulated noise. *The Journal of the Acoustical Society of America*, *70*(6), 1655–1660.

Buss, E., Lorenzi, C., Cabrera, L., Leibold, L. J., and Grose, J. H. (2019). Amplitude modulation detection and modulation masking in school-age children and adults. *The Journal of the Acoustical Society of America*, *145*(4), 2565–2575.

Byrne, A. J., Stellmack, M. A., and Viemeister, N. F. (2013). The salience of enhanced components within inharmonic complexes. *The Journal of the Acoustical Society of America*, *134*(4), 2631–2634.

Cao, X., and Richards, V. M. (2012). Enhancement in informational masking. *Journal of Speech, Language, and Hearing Research*, *55*, 1135–1147.

Conroy, C., and Kidd Jr, G. (2021). Informational masking in the modulation domain. *The Journal of the Acoustical Society of America*, *149*(5), 3665–3673.

Dau, T., Kollmeier, B., and Kohlrausch, A. (1997). Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers. *The Journal of the Acoustical Society of America*, *102*(5), 2892–2905.

Dau, T., Verhey, J., and Kohlrausch, A. (1999). Intrinsic envelope fluctuations and modulation-detection thresholds for narrow-band noise carriers. *The Journal of the Acoustical Society of America, 106*(5), 2752–2760.

Drullman, R., Festen, J. M., and Plomp, R. (1994a). Effect of temporal envelope smearing on speech reception. *The Journal of the Acoustical Society of America*, *95*(2), 1053–1064.

Drullman, R., Festen, J. M., and Plomp, R. (1994b). Effect of reducing slow temporal modulations on speech reception. *The Journal of the Acoustical Society of America*, *95*(5), 2670–2680.

Dubbelboer, F., and Houtgast, T. (2008). The concept of signal-to-noise ratio in the modulation domain and speech intelligibility. *The Journal of the Acoustical Society of America, 124*(6), 3937–3946.

Durlach, N. I., and Braida, L. D. (1969). Intensity perception. I. Preliminary theory of intensity resolution. *The Journal of the Acoustical Society of America*, *46*(2B), 372–383.

Durlach, N. I., Mason, C. R., Kidd Jr, G., Arbogast, T. L., Colburn, H. S., and Shinn-Cunningham, B. G. (2003a). Note on informational masking (L). *The Journal of the Acoustical Society of America*, *113*(6), 2984–2987.

Durlach, N. I., Mason, C. R., Shinn-Cunningham, B. G., Arbogast, T. L., Colburn, H. S., and Kidd Jr, G. (2003b). Informational masking: Counteracting the effects of stimulus uncertainty by decreasing target-masker similarity. *The Journal of the Acoustical Society of America*, *114*(1), 368–379.

Durlach, N. I., Mason, C. R., Gallun, F. J., Shinn-Cunningham, B., Colburn, H. S., and Kidd Jr, G. (2005). Informational masking for simultaneous nonspeech stimuli: Psychometric functions for fixed and randomly mixed maskers. *The Journal of the Acoustical Society of America*, *118*(4), 2482–2497.

Ewert, S. D., and Dau, T. (2000). Characterizing frequency selectivity for envelope fluctuations. *The Journal of the Acoustical Society of America*, *108*(3), 1181–1196.

Ewert, S. D., Verhey, J. L., and Dau, T. (2002). Spectro-temporal processing in the envelope-frequency domain. *The Journal of the Acoustical Society of America*, *112*(6), 2921–2931.

Fan, W. L., Streeter, T. M., and Durlach, N. I. (2008). Effect of spatial uncertainty of masker on masked detection for nonspeech stimuli. *The Journal of the Acoustical Society of America*, *124*(1), 36–39.

Fastl, H. (1983). Fluctuation strength of modulated tones and broadband noise. In R. Klinke and R. Hartmann (Eds.), *Hearing – physiological bases and psychophysics* (pp. 282–288). Springer.

Feng, L., and Oxenham, A. J. (2015). New perspectives on the measurement and time course of auditory enhancement. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(6), 1696.

Füllgrabe, C., Moore, B. C. J., Demany, L., Ewert, S. D., Sheft, S., and Lorenzi, C. (2005). Modulation masking produced by second-order modulators. *The Journal of the Acoustical Society of America*, *117*(4), 2158–2168.

Füllgrabe, C., Sęk, A., and Moore, B. C. J. (2021). Forward masking of amplitude modulation across ears and its tuning in the modulation domain. *The Journal of the Acoustical Society of America*, *149*(3), 1764–1771.

Fu, Q. J. (2002). Temporal processing and speech recognition in cochlear implant users. *NeuroReport*, *13*(13), 1635–1639.

Green, D. M. (1995). Maximum-likelihood procedures and the inattentive observer. *The Journal of the Acoustical Society of America*, *97*(6), 3749–3760.

Grimault, N., Bacon, S. P., and Micheyl, C. (2002). Auditory stream segregation on the basis of amplitude-modulation rate. *The Journal of the Acoustical Society of America*, *111*(3), 1340–1348.

Hanna, T. E. (1992). Discrimination and identification of modulation rate using a noise carrier. *The Journal of the Acoustical Society of America*, *91*(4), 2122–2128.

Hautus, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated values of d'. *Behavior Research Methods, Instruments, & Computers*, *27*(1), 46–51.

Houtgast, T. (1989). Frequency selectivity in amplitude-modulation detection. *The Journal of the Acoustical Society of America*, *85*(4), 1676–1680.

Hübner, R. (1993). On possible models of attention in signal detection. *Journal of Mathematical Psychology*, *37*(2), 266–281.

Huang, R., and Richards, V. M. (2006). Coherence detection: Effects of frequency, frequency uncertainty, and onset⁄ offset delays. *The Journal of the Acoustical Society of America*, *119*(4), 2298–2304.

Huang, R., and Richards, V. M. (2008). Estimates of internal templates for the detection of sequential tonal patterns. *The Journal of the Acoustical Society of America*, *124*(6), 3831–3840.

Jørgensen, S., and Dau, T. (2011). Predicting speech intelligibility based on the signal-to-noise envelope power ratio after modulation-frequency selective processing. *The Journal of the Acoustical Society of America*, *130*(3), 1475–1487.

Joris, P. X., Schreiner, C. E., and Rees, A. (2004). Neural processing of amplitude-modulated sounds. *Physiological Reviews*, *84*(2), 541–577.

Kidd Jr, G., Mason, C. R., Deliwala, P. S., Woods, W. S., and Colburn, H. S. (1994). Reducing informational masking by sound segregation. *The Journal of the Acoustical Society of America*, *95*(6), 3475–3480.

Kidd Jr, G., Mason, C. R., and Dai, H. (1995). Discriminating coherence in spectro-temporal patterns. *The Journal of the Acoustical Society of America*, *97*(6), 3782–3790.

Kidd Jr, G., Mason, C. R., and Arbogast, T. L. (2002). Similarity, uncertainty, and masking in the identification of nonspeech auditory patterns. *The Journal of the Acoustical Society of America*, *111*(3), 1367–1376.

Kidd Jr, G., Mason, C. R., and Richards, V. M. (2003). Multiple bursts, multiple looks, and stream coherence in the release from informational masking. *The Journal of the Acoustical Society of America*, *114*(5), 2835–2845.

Kidd, G., Mason, C. R., Richards, V. M., Gallun, F. J., and Durlach, N. I. (2008). Informational masking. In W. A. Yost, A. N. Popper, and R. R. Fay (Eds.), *Auditory perception of sound sources, Springer handbook of auditory research*, *29* (pp. 143–189). Springer.

Kidd Jr, G., Richards, V. M., Streeter, T., Mason, C. R., and Huang, R. (2011). Contextual effects in the identification of nonspeech auditory patterns. *The Journal of the Acoustical Society of America*, *130*(6), 3926–3938.

Kohlrausch, A., Fassel, R., and Dau, T. (2000). The influence of carrier level and frequency on modulation and beat-detection thresholds for sinusoidal carriers. *The Journal of the Acoustical Society of America*, *108*(2), 723–734.

Langner, G., and Schreiner, C. E. (1988). Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms. *Journal of Neurophysiology*, *60*(6), 1799–1822.

Lee, T. Y., and Richards, V. M. (2011). Evaluation of similarity effects in informational masking. *The Journal of the Acoustical Society of America*, *129*(6), EL280–EL285.

Leibold, L. J., Hitchens, J. J., Buss, E., and Neff, D. L. (2010). Excitation-based and informational masking of a tonal signal in a four-tone masker. *The Journal of the Acoustical Society of America*, *127*(4), 2441–2450.

Levitt, H. C. (1971). Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical society of America*, *49*(2B), 467–477.

Litovsky, R. Y., Parkinson, A., and Arcaroli, J. (2009). Spatial hearing and speech intelligibility in bilateral cochlear implant users. *Ear and Hearing*, *30*(4), 419.

Lorenzi, C., Soares, C., and Vonner, T. (2001). Second-order temporal modulation transfer functions. *The Journal of the Acoustical Society of America*, *110*(2), 1030–1038.

Lutfi, R. A. (1990). How much masking is informational masking? *The Journal of the Acoustical Society of America*, *88*(6), 2607–2610.

Lutfi, R. A. (1993). A model of auditory pattern analysis based on component-relative-entropy. *The Journal of the Acoustical Society of America*, *94*(2), 748–758.

Lutfi, R. A., Kistler, D. J., Oh, E. L., Wightman, F. L., and Callahan, M. R. (2003a). One factor underlies individual differences in auditory informational masking within and across age groups. *Perception & Psychophysics*, *65*(3), 396–406.

Lutfi, R. A., Kistler, D. J., Callahan, M. R., and Wightman, F. L. (2003b). Psychometric functions for informational masking. *The Journal of the Acoustical Society of America*, *114*(6), 3273–3282.

Macmillan, N. A., and Creelman, C. D. (1990). Response bias: Characteristics of detection theory, threshold theory, and "nonparametric" indexes. *Psychological Bulletin*, *107*(3), 401.

Macmillan, N. A., and Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Lawrence Erlbaum Associates.

McDermott, J. H., and Simoncelli, E. P. (2011). Sound texture perception via statistics of the auditory periphery: evidence from sound synthesis. *Neuron*, *71*(5), 926–940.

McDermott, J. H., Schemitsch, M., and Simoncelli, E. P. (2013). Summary statistics in auditory perception. *Nature Neuroscience*, *16*(4), 493–498.

McWalter, R., and Dau, T. (2017). Cascaded amplitude modulations in sound texture perception. *Frontiers in Neuroscience*, *11*, 485.

Micheyl, C., Shamma, S.A., and Oxenham, A.J. (2007). Hearing out repeating elements in randomly varying multitone sequences: a case of streaming? In B. Kollmeier, G. Klump, V. Hohmann, U. Langemann, M. Mauermann, S. Uppenkamp, and J. Verhey (Eds.), *Hearing – from basic research to application* (pp. 267–274). Springer.

Micheyl, C., Kreft, H., Shamma, S., and Oxenham, A. J. (2013). Temporal coherence versus harmonicity in auditory stream formation. *The Journal of the Acoustical Society of America*, *133*(3), EL188–EL194.

Moore, B. C. J. (1995). Frequency analysis and masking. In B. C. J. Moore (Ed.), *Hearing*, (pp. 161–205). Academic.

Moore, B. C. J., Füllgrabe, C., and Sek, A. (2009). Estimation of the center frequency of the highest modulation filter. *The Journal of the Acoustical Society of America*, *125*(2), 1075–1081.

Neff, D. L., and Green, D. M. (1987). Masking produced by spectral uncertainty with multicomponent maskers. *Perception & Psychophysics, 41*(5), 409–415.

Neff, D. L., and Callaghan, B. P. (1987). Psychometric functions for multicomponent maskers with spectral uncertainty. *The Journal of the Acoustical Society of America*, *81*(S1), S53–S53.

Neff, D. L., and Callaghan, B. P. (1988). Effective properties of multicomponent simultaneous maskers under conditions of uncertainty. *The Journal of the Acoustical Society of America*, *83*(5), 1833–1838.

Neff, D. L., Dethlefs, T. M., and Jesteadt, W. (1993). Informational masking for multicomponent maskers with spectral gaps. *The Journal of the Acoustical Society of America*, *94*(6), 3112–3126.

Neff, D. L., and Dethlefs, T. M. (1995). Individual differences in simultaneous masking with random-frequency, multicomponent maskers. *The Journal of the Acoustical Society of America*, *98*(1), 125–134.

Neff, D. L. (1995). Signal properties that reduce masking by simultaneous, random-frequency maskers. *The Journal of the Acoustical Society of America*, *98*(4), 1909–1920.

Oh, E. L., and Lutfi, R. A. (1998). Nonmonotonicity of informational masking. *The Journal of the Acoustical Society of America*, *104*(6), 3489–3499.

Oxenham, A. J., Fligor, B. J., Mason, C. R., and Kidd Jr, G. (2003). Informational masking and musical training. *The Journal of the Acoustical Society of America*, *114*(3), 1543–1549.

Pollack, I. (1975). Auditory informational masking. *The Journal of the Acoustical Society of America*, *57*(S1), S5–S5.

Richards, V. M., Buss, E., and Tian, L. (1997). Effects of modulator phase for comodulation masking release and modulation detection interference. *The Journal of the Acoustical Society of America*, *102*(1), 468–476.

Richards, V. M., Buss, E., and Tian, L. (1998). The effects of different envelope patterns and uncertainty for the detection of a tone added to SAM complex tonal maskers. *The Journal of the Acoustical Society of America*, *103*(2), 1058–1066.

Richards, V. M., Tang, Z., and Kidd Jr, G. D. (2002). Informational masking with small set sizes. *The Journal of the Acoustical Society of America*, *111*(3), 1359–1366.

Richards, V. M., and Neff, D. L. (2004). Cuing effects for informational masking. *The Journal of the Acoustical Society of America*, *115*(1), 289–300.

Richards, V. M., Huang, R., and Kidd Jr, G. (2004). Masker-first advantage for cues in informational masking. *The Journal of the Acoustical Society of America*, *116*(4), 2278–2288.

Richards, V. M., and Tang, Z. (2006). Estimates of effective frequency selectivity based on the detection of a tone added to complex maskers. *The Journal of the Acoustical Society of America*, *119*(3), 1574–1584.

Richards, V. M., Tisby, M. K., Suzuki-Gill, E. N., and Shen, Y. (2021). Sub-optimal construction of an auditory profile from temporally distributed spectral information. *The Journal of the Acoustical Society of America*, *149*(3), 1567–1578.

Ritsma, R. J. (1962). Existence region of the tonal residue. *I. The Journal of the Acoustical Society of America*, *34*(9A), 1224–1229.

Sek, A., and Moore, B. C. J. (2003). Testing the concept of a modulation filter bank: The audibility of component modulation and detection of phase change in three-component modulators. *The Journal of the Acoustical Society of America*, *113*(5), 2801–2811.

Sek, A., Baer, T., Crinnion, W., Springgay, A., and Moore, B. C. J. (2015). Modulation masking within and across carriers for subjects with normal and impaired hearing. *The Journal of the Acoustical Society of America*, *138*(2), 1143–1153.

Shamma, S. A., Elhilali, M., and Micheyl, C. (2011). Temporal coherence and attention in auditory scene analysis. *Trends in Neurosciences*, *34*(3), 114–123.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, *270*(5234), 303–304.

Sheft, S., and Yost, W. A. (2007). Modulation detection interference as informational masking. In B. Kollmeier, G. Klump, V. Hohmann, U. Langemann, M. Mauermann, S. Uppenkamp, and J. Verhey (Eds.), *Hearing – from sensory processing to perception* (pp. 303–311). Springer.

Shen, Y. (2017). Auditory sequential accumulation of spectral information. *Hearing Research*, *356*, 118–126.

Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends in Cognitive Sciences*, *12*(5), 182–186.

Simpson, A. J., and Fitter, M. J. (1973). What is the best index of detectability? *Psychological Bulletin*, *80*(6), 481.

Spiegel, M. F., and Green, D. M. (1982). Signal and masker uncertainty with noise maskers of varying duration, bandwidth, and center frequency. *The Journal of the Acoustical Society of America*, *71*(5), 1204–1210.

Stone, M. A., Füllgrabe, C., and Moore, B. C. J. (2012). Notionally steady background noise acts primarily as a modulation masker of speech. *The Journal of the Acoustical Society of America*, *132*(1), 317–326.

Stone, M. A., and Moore, B. C. J. (2014). On the near non-existence of "pure" energetic masking release for speech. *The Journal of the Acoustical Society of America*, *135*(4), 1967–1977.

Strickland, E. A., and Viemeister, N. F. (1996). Cues for discrimination of envelopes. *The Journal of the Acoustical Society of America*, *99*(6), 3638–3646.

Tang, Z., and Richards, V. M. (2003). Examination of a linear model in an informational masking study. *The Journal of the Acoustical Society of America*, *114*(1), 361–367.

Tansley, B. W., and Suffield, J. B. (1983). Time course of adaptation and recovery of channels selectively sensitive to frequency and amplitude modulation. *The Journal of the Acoustical Society of America*, *74*(3), 765–775.

Tanner, W. P., Jr. (1961). Physiological implications of psychophysical data. *Annals of the New York Academy of Sciences*, *89*, 752–765.

Tansley, B. W., and Suffield, J. B. (1983). Time course of adaptation and recovery of channels selectively sensitive to frequency and amplitude modulation. *The Journal of the Acoustical Society of America*, *74*(3), 765–775.

Teki, S., Chait, M., Kumar, S., Shamma, S., and Griffiths, T. D. (2013). Segregation of complex acoustic scenes based on temporal coherence. *elife*, *2*, e00699.

Terhardt, E. (1968). Über akustische Rauhigkeit und Schwankungsstärke (On the acoustic roughness and fluctuation strength). *Acustica*, *20*, 215–224.

Van Tasell, D. J., Soli, S. D., Kirby, V. M., and Widin, G. P. (1987). Speech waveform envelope cues for consonant recognition. *The Journal of the Acoustical Society of America*, *82*(4), 1152–1161.

Viemeister, N. (1980). Adaptation of masking. In G. van den Brink and F. A. Bilsen (Eds.), *Psychophysical, physiological and behavioral studies in hearing,* (pp. 190–198). Delft University Press.

Viemeister, N. F., and Bacon, S. P. (1982). Forward masking by enhanced components in harmonic complexes. *The Journal of the Acoustical Society of America*, *71*(6), 1502–1507.

Viemeister N.F., Stellmack M.A., and Byrne A.J. (2005). The role of temporal structure in envelope processing. In D. Pressnitzer, A. de Cheveigne, S. McAdams, and L. Collet L. (Eds.), *Auditory Signal Processing* (pp. 220–228). Springer.

Watson, C. S., and Kelly, W. J. (1981). The role of stimulus uncertainty in the discriminability of auditory patterns. In D. J. Getty and J. H. Howard, Jr. (Eds.), *Auditory and visual pattern recognition* (pp. 37–59). Erlbaum.

Watson, C. S. (1987). Uncertainty, informational masking and the capacity of immediate auditory memory. In W.A. Yost and C.S. Watson (Eds.), *Auditory processing of complex sounds* (pp. 267–277). Erlbaum.

Watson, C. S. (2005). Some comments on informational masking. *Acta Acustica united with Acustica*, *91*(3), 502–512.

Wojtczak, M., and Viemeister, N. F. (2003). Suprathreshold effects of adaptation produced by amplitude modulation. *The Journal of the Acoustical Society of America*, *114*(2), 991–997.

Wojtczak, M., and Viemeister, N. F. (2005). Forward masking of amplitude modulation: Basic characteristics. *The Journal of the Acoustical Society of America*, *118*(5), 3198–3210.

Wojtczak, M., and Viemeister, N. F. (2008). Perception of suprathreshold amplitude modulation and intensity increments: Weber's law revisited. *The Journal of the Acoustical Society of America*, *123*(4), 2220–2236.

Wojtczak, M., Nelson, P. C., Viemeister, N. F., and Carney, L. H. (2011). Forward masking in the amplitude-modulation domain for tone carriers: psychophysical results and physiological correlates. *Journal of the Association for Research in Otolaryngology*, *12*(3), 361–373.

Wojtczak, M. (2011). The effect of carrier level on tuning in amplitude-modulation masking. *The Journal of the Acoustical Society of America*, *130*(6), 3916–3925.

Wright, B. A., and McFadden, D. (1990). Uncertainty about the correlation among temporal envelopes in two comodulation tasks. *The Journal of the Acoustical Society of America*, *88*(3), 1339–1350.

Wright, B. A., and Saberi, K. (1999). Strategies used to detect auditory signals in small sets of random maskers. *The Journal of the Acoustical Society of America*, *105*(3), 1765–1775.

CURRICULUM VITAE

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]