October 2021

THE BENEFITS OF SPATIAL SEPARATION ON THE CORTICAL REPRESENTATIONS OF SPEECH SOUNDS

Benjamin H. Zobel
University of Massachusetts Amherst

THE BENEFITS OF SPATIAL SEPARATION ON THE CORTICAL REPRESENTATIONS OF SPEECH SOUNDS

Benjamin H. Zobel

THE BENEFITS OF SPATIAL SEPARATION ON THE CORTICAL REPRESENTATIONS OF SPEECH SOUNDS

A Dissertation Presented

by

BENJAMIN H. ZOBEL

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

September 2021

Psychological and Brain Sciences

THE BENEFITS OF SPATIAL SEPARATION ON THE CORTICAL REPRESENTATIONS OF SPEECH SOUNDS

A Dissertation Presented

by

BENJAMIN H. ZOBEL

Approved as to style and content by:

_____
Lisa D. Sanders, Chair

_____
Richard L. Freyman, Member

_____
Charles Clifton, Jr., Member

_____
Neil E. Berthier, Member

_____
Farshid Hajir, Department Head
Psychological and Brain Sciences

DEDICATION

To Kat, Phineas, and Grey

ACKNOWLEDGMENTS

I thank my advisor, Lisa Sanders, for her mentorship, guidance, support and patience throughout my graduate career. I owe the bulk of my education and success to her immense skills as a supervisor, teacher, and collaborator, and her brilliance as a research scientist.

I thank my committee members, Rich Freyman, Chuck Clifton, and Neil Berthier, for their valuable support and feedback throughout the dissertation process and throughout my graduate career.

I thank the students within the Neurocognition and Perception Lab and the faculty within the Cognition and Cognitive Neuroscience division at the University of Massachusetts Amherst for the privilege to work and to learn within a collaborative, engaging, and rewarding research environment.

I would also like to thank a set of individuals whose contributions are in one way or another reflected in the work presented here, either by assisting with the preparation and running of the studies, or showing me the ropes and getting me up to speed when I first entered the graduate program, or providing me with guidance, comradery, and an inspiring exchange of ideas throughout my graduate career: Hillary Hadley, Maggie Ugolini, Amanda Rysling, Jen Bieber, Miriam Munoz, Lori Astheimer Best, Dave Katz, Amy Joh, Nick Planet, Will Bush, Ahren Fitzroy, Chad Dubé, and Tony McCaffrey.

Finally, I thank my parents, Fred and Teri, for their continuous support through the years.

ABSTRACT

THE BENEFITS OF SPATIAL SEPARATION ON THE CORTICAL REPRESENTATIONS OF SPEECH SOUNDS

SEPTEMBER 2021

BENJAMIN H. ZOBEL, B.A., NEW YORK UNIVERSITY

M.A., UNIVERSITY OF MASSACHUSETTS AMHERST

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Lisa D. Sanders

Spatial separation between competing speech streams reduces their confusion (informational masking) and improves speech processing under challenging listening conditions. The precise stages of auditory processing and the bottom-up and top-down mechanisms involved in this spatial release from informational masking are not fully understood. Event-related potentials (ERPs) were used to measure the cortical processing of relevant speech under conditions of informational masking and its spatial release, and to examine the preattentive and attentive mechanisms that benefit listeners.

Participants were asked to detect noise-vocoded target speech presented with noise-vocoded two-talker masking speech. In separate conditions, the same set of targets were spatially co-located with maskers to produce a high degree of informational masking and spatially separated from maskers using a perceptual manipulation to release the informational masking. Cortical auditory evoked potentials (N1/P2 ERP waveforms) elicited by targets were only apparent under conditions in which informational masking was released. Furthermore, when targets were presented at an intensity above masking threshold in both spatial conditions, N1 and P2 latencies were shorter when targets were perceptually separated compared to co-located with maskers. These effects of spatial

separation were observed regardless of whether participants attended to the auditory task or attended away from the sounds to engage in a challenging two-back visual task. Benefits of attending to the sounds were apparent in later time windows (P2 and P3), while there was tentative evidence of attentional benefits earlier in processing under some conditions.

These results show that spatial separation between competing speech streams can facilitate preattentive, bottom-up processes that reduce confusion and improve the early perceptual representations of relevant speech. Top-down selective attention is necessary for supporting higher-level task-relevant cognitive benefits of spatial separation that occur at later stages of processing, and may play a more crucial role in the early perceptual processing of speech under especially challenging listening conditions. These studies shed light on the underlying processes that contribute to the spatial release from informational masking, and establish methods and measures that may be applied in future research aiming to benefit listeners who experience difficulties processing speech within noisy, complex environments.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

CHAPTER 1

STUDY 1: THE ERP INDICES OF SPATIAL RELEASE FROM INFORMATIONAL MASKING

1.1 Introduction

Listeners are commonly faced with the challenge of processing speech within noisy, complex environments. Often, this challenge takes the form of a "cocktail party problem" in which a listener must understand what one person is saying while others around them are talking (Cherry, 1953). Listeners benefit when the source of the relevant speech (target) and the source of the interfering noise (masker) are spatially separated (Bronkhorst, 2000, 2015). This benefit, commonly called spatial release from masking, is behaviorally measured as a decrease in the threshold for target detection or speech identification (recognition, comprehension, etc.) when target and masker are presented from separate spatial locations compared to the same spatial location. Since not all listeners benefit equally from the spatial separation available in typical real-world listening conditions (Arbogast et al., 2005; Zobel et al., 2019), it is important to understand the mechanisms that support improved speech processing with spatial separation.

Research defines two broad categories of masking—energetic and informational masking. Both types of masking are present at a typical cocktail party, and both can be released by spatially separating target and masker (Freyman et al., 1999; Zurek, 1993). Energetic masking occurs at the initial stages of processing when the signal-to-noise ratio (SNR; the intensity of the target relative to the intensity of the masker) is sufficiently low and the degree of spectral overlap between target and masker sufficiently high that the

masker energy saturates the sensory resources needed for target encoding (Fletcher, 1940; Miller, 1947). Moving target and masker into separate spatial locations can produce head shadow effects (Shaw, 1974) and binaural interactions (Licklider, 1948) that reduce energetic masking. Reductions in energetic masking can account for the spatial release from masking observed under simple anechoic conditions (Zurek, 1993). Everyday environments, however, are far more complex, often involving multiple sound sources and reflections that arrive at the ears from different directions that may swamp out these sensory benefits; however, benefits of spatial separation are still observed that are attributed to a release from informational masking (Freyman et al., 1999; Kidd, Mason, et al., 2005).

Informational masking describes a confusion between the target and masker that arises when there is a lack of distinguishing cues between the two. Figure 1 (Lutfi et al., 2013) provides a visual analog of informational masking. In these images, the target information (Figure 1a) is clearly displayed without obstruction from view (i.e., no energetic masking). However, perceptual similarity and lack of a predictable spatial pattern results in informational masking (Figure 1b) that is released when target and masker are perceptually dissimilar and predictable (Figure 1c). Informational masking in the auditory modality involves a similar target/masker confusion that produces masking in excess of what can be accounted for by energetic masking (Kidd et al., 2008). For example, listeners can experience greater masking when a speech target is presented with a two-talker speech masker compared to a broad-band steady-state noise masker (Carhart et al., 1969). Energetic masking is lower with the speech masker, which includes dips in intensity that allow for better sensory encoding of the target, but informational masking

can be so high with the speech masker that it more than overwhelms the benefits of reduced energetic masking (Carhart et al., 1969). Targets and maskers that are more similar to each other result in more target/masker confusion and more informational masking (Brungart et al., 2001; Festen & Plomp, 1990). Listeners have been shown to take advantage of almost any cue that distinguishes targets from maskers, from the features of a voice and statistical regularities in simple non-speech stimuli to the content of longer passages of speech (Başkent & Gaudrain, 2016; Bradlow & Alexander, 2007; Cherry, 1953; Darwin & Hukin, 2000; Freyman et al., 2004; Kidd et al., 1994; Watson et al., 1975). The wide range of perceptual and cognitive levels of processing at which informational masking can be released has contributed to a lack of precision in defining what informational masking is (Durlach et al., 2003; Kidd et al., 2008; Watson, 2005) and why some populations struggle to find release from informational masking in the noisy complex environments of everyday life (Arbogast et al., 2005; Zobel et al., 2019). To make progress towards understanding the mechanisms that support release from informational masking, it is important to isolate informational masking and a single source of release from that masking to the greatest extent possible.

The strength with which informational masking can be released by a spatial cue is particularly evident under conditions of virtual spatial separation (Freyman et al., 1999). In this paradigm, one loudspeaker is placed in front (F) of a listener and another loudspeaker is placed to the listener's right (R). In the spatially co-located condition both the target and masker are presented from the front loudspeaker while no stimulus is presented from the right loudspeaker (F-F condition). In the virtual spatially separated condition, target and masker are again both presented from the front loudspeaker while an

identical copy of the masker is presented from the right with its onset preceding the front masker by 4 ms (F-RF condition). This 4-ms right-lead stimulus onset asynchrony (SOA) creates the precedence effect in listeners, a mechanism for accurate source localization within reverberant environments that fuses direct and reflected sounds into a single auditory object localized at the source of the direct sound (Litovsky et al., 1999). As a result, listeners perceive only a single masker presented from the right that is now spatially separated from the front target, despite the fact that the target and masker at the front loudspeaker remain physically co-located across conditions. Release from speech-on-speech masking can be large in the F-RF condition, with a threshold reduction upwards of 20 dB for the detection of natural or vocoded target speech compared to the F-F condition (Brungart et al., 2005; Freyman et al., 1999, 2001, 2004, 2008; Morse-Fortier et al., 2017; Rakerd et al., 2006; X. Wu et al., 2005; Zobel et al., 2019). Importantly, evidence suggests that only informational masking is released in the F-RF condition, to the extent that it is present in the F-F condition. Masking release is not observed if target and masker are perceptually distinct (e.g., speech or narrow-band noise target with broadband noise masker), suggesting that energetic masking remains minimally affected by the spatial manipulation (Brungart et al., 2005; Freyman et al., 1999, 2004, 2008; Morse-Fortier et al., 2017; Rakerd et al., 2006). Thus, the virtual separation paradigm provides the desired isolation of informational masking. There is evidence that the spatial cue facilitates both the bottom-up grouping of target and masker into independent auditory objects and the top-down allocation of attention to the target (Ihlefeld & Shinn-Cunningham, 2008b, 2008a; Zhang et al., 2014, 2019), but questions remain about the underlying mechanisms and their relative contributions.

To examine the stages of processing involved in the spatial release from informational masking, the present study used event-related potentials (ERPs) to measure responses to targets presented within the virtual separation paradigm. Two recent studies have used ERPs to examine spatial release from masking under virtual separation (Zhang et al., 2014, 2019). Both studies suggest that a spatial cue may facilitate early perceptual representations of target speech as indexed by an amplitude increase and latency decrease in the cortical auditory evoked potentials (N1 and P2 waveforms) elicited by a target syllable when virtually separated from a two-talker speech masker. However, behavioral evidence suggests that the targets in these studies were presented at an SNR that was well above masking threshold in both the spatially co-located and virtually separated conditions (Zhang et al., 2016, 2019). Therefore, the extent to which masking was present and spatially released in these studies remains unclear. Moreover, the ERP effects of introducing the spatial cue in the absence of other distinguishing cues in natural speech that were likely to benefit listeners in these studies (e.g., pitch differences among talkers) remains unknown. The present study minimized all distinguishing cues in the stimuli except for the spatial cue, and measured ERPs elicited by the same targets when they were masked in the F-F condition and unmasked in the F-RF condition. A detection task was used to establish thresholds that are dependent on the perceptual separation of target and masker. Targets were noise-vocoded single-syllable words (female talker) with sharp onsets for eliciting cortical auditory evoked potentials. Maskers were noise-vocoded segments of two-talker babble (separate female talkers). Participants were asked to detect whether or not a target was present on each trial with a "yes" or "no" response. To minimize distinguishing cues and maximize informational masking, the targets and

maskers were noise-vocoded to increase target/masker similarity, and the content of targets and timing of their onsets varied from trial to trial to increase uncertainty. Detection thresholds for this type of stimuli in the F-F condition consistently approach 0 dB SNR (Freyman et al., 2008; Morse-Fortier et al., 2017; Zobel et al., 2019), the hypothesized limit for informational masking (Arbogast et al., 2005; Freyman et al., 2008). Thus, target/masker confusion is maximized by a lack of cues in the F-F condition, and the large release from masking that is typically observed in the F-RF condition can be solely attributed to the spatial cue.

Prior to beginning the experiment, a screening procedure confirmed that every participant who contributed data for analysis exhibited a large release from masking. The screening was also used to choose for each participant an ideal set of SNRs at which to present targets for examining the ERP effects of spatial separation. In addition to SNRs below and above masking threshold in both spatial conditions, a crucial SNR was chosen at which targets would be masked in the F-F condition and unmasked in the F-RF condition, thus capturing the effects of spatial release from masking. As a result, behavioral responses and ERPs time-locked to identical sets of targets presented at these SNRs provided a comprehensive account of the extent to which spatial separation alone serves to reduce target/masker confusion and improve target representation at different stages of processing under challenging listening conditions.

1.2 Methods

The study consisted of two sections conducted in one ~2.5-hr session: 1) a screening section, in which hearing was assessed with pure-tone audiometry, and behavioral data from the target-detection task was collected to evaluate masking

thresholds and determine the stimuli to be used in the experiment, and 2) an experimental section, in which behavioral and EEG data were simultaneously recorded while participants completed the target detection task with the chosen stimuli in the F-F and F-RF conditions.

1.2.1 Participants

Twenty right-handed, native English-speaking participants reporting no known neurological problems or use of psychoactive medication at the time of the study contributed the behavioral and ERP data for analysis (9 female, 19 – 33 years, $M = 24.7$ years, $SD = 4.45$ years). An additional nine participants did not complete the experiment, because either their preliminary hearing assessment exceeded +20 dB hearing level at either 1, 2, 4, or 8 kHz ($n = 3$), thresholds in the F-F and F-RF conditions did not show spatial release from masking (n = 2), or thresholds could not be determined from behavioral responses in the screening section with sufficient consistency to select the experimental stimuli ($n = 4$). All participants provided informed consent and were compensated at a rate of $10/hr.

1.2.2 Stimuli

1.2.2.1 Target Stimuli

Targets were single-syllable noise-vocoded words. Target words consisted of 80 screening words (used in the screening section) and 80 experimental words (used in the experimental section) chosen from the unrestricted lexicon of the English Lexicon Project (Balota et al., 2007). Criteria were applied for selecting target words with sharp acoustic onsets to elicit auditory evoked potentials, and lower phonological-neighbor frequencies

to reduce any influence of such cues in target detection: 1) Target words had to be one syllable long, 2) target words had to begin with a stop consonant ([b], [d], [g], [k], [p], [t]), 3) target-word frequency had to be below the *SUBTLEX* average frequency of the lexicon, 4) the number of phonological neighbors of a target word could not exceed 20 for the screening words and 12 for the experimental words, and 5) target words could not have an extremely frequent neighbor (e.g., a neighbor like *can* or *go*). To promote consistency and natural articulation, a female voice was recorded reading each target word embedded at the end of a sentence in the following form: *The next word is [target word]*. Target words were then extracted from their respective sentences with Pro Tools audio software (Avid Technology) and processed in MATLAB (The MathWorks Inc.) with a six-channel white-noise vocoder used in prior speech-masking studies (Freyman et al., 2008; Morse-Fortier et al., 2017; Qin & Oxenham, 2003; Zobel et al., 2019).

1.2.2.2 Masker Stimuli

Maskers were two-talker noise-vocoded female babble. Two females, different from the target talker, were recorded reading a list of 900 semantically nonsensical sentences (e.g., *His throat could knit the coward*) developed for similar speech-masking studies (Freyman et al., 1999, 2004, 2007, 2008). Each female voice was recorded in isolation reading the sentences one after the other. Recordings were edited to remove errors and limit the duration of gaps in speech to no longer than 100 ms, resulting in two continuous speech streams, each ~30 min in length. Each stream was then divided into 960 individual 2500-ms segments. The segments were each vocoded and scaled to the same root mean square (RMS) amplitude. Each vocoded segment from one female talker was then paired with a randomly selected segment from the other female talker under the

constraint that the paired segments could not overlap in their content. Each pair was summed together and then scaled to the same RMS amplitude, resulting in 960 different 2500-ms vocoded two-talker maskers.

1.2.3 Stimulus Conditions

Fifty-one copies of each target were created and scaled in 1-dB steps such that their RMS amplitudes ranged from -40 dB to +10 dB relative to the Masker RMS amplitude. Each target was saved as a stereo WAV file (44.1 kHz sampling rate, 16-bit resolution) with the target placed in channel 1 and silence in channel 2. Two spatial versions of each masker were created and saved in separate stereo WAV files. The F-F maskers were created by placing a masker in channel 1 and silence in channel 2. The F-RF maskers were created by placing the same masker in both channels with a 4-ms SOA between them (channel 2 preceding channel 1). Note that target intensity was varied while masker intensity was held constant throughout the study, and the reported SNRs in both spatial conditions were calculated as the RMS amplitude of the target relative to the masker in channel 1.

1.2.4 Study Setup

Participants were seated in the center of a sound-dampened 2.5-m x 3.5-m room with a matched pair of magnetically shielded Genelec 8030A loudspeakers placed 1.5 m in front of them at 0° and 55° right of midline, respectively. A computer monitor was positioned directly beneath the front loudspeaker to display text (e.g., instructions, fixation cross) in a white font against a black background. The auditory stimuli were presented with E-Prime software (Psychology Software Tools, Inc.) running on a PC

equipped with a Creative Labs Sound Blaster Audigy 2 ZS sound card. The digital audio signal was sent to an M-Audio Super DAC 2496 D/A converter outputting channel 1 to the front loudspeaker and channel 2 to the right loudspeaker. Loudspeaker intensities were equated prior to beginning the study by separately adjusting their gains until they each individually presented maskers at 70 dBA SPL at the participant position.

1.2.5 Procedure

1.2.5.1 Screening Section

The screening was designed to measure F-F and F-RF masking thresholds and determine the SNRs to be used in the experiment according to the criteria described in Table 1. These experimental SNRs included ones below and above masking threshold in both spatial conditions (SNR_{LOW} and SNR_{HIGH}), an SNR at which targets were masked in the F-F condition and unmasked in the F-RF condition to best capture spatial release from masking (SNR_{SRM}), and SNRs 2 dB above and below SNR_{SRM} in case responses at SNR_{SRM} ended up deviating from the criteria for some participants.

Following their hearing assessment, participants were familiarized with the stimuli by listening to several targets presented in isolation. The targets were described as examples of a particular "target voice" that, when present on a trial, would always come from the front loudspeaker and say only a single word. Participants were told that some trials would contain the target voice and some would not, and their task was to detect whether or not the target voice was present among other "babbling voices" that would come from either the front or right loudspeaker. To exemplify these babbling voices, several instances of the F-F and F-RF maskers were presented in isolation.

Targets with high SNRs (+8 to +10 dB) were then presented with F-F and F-RF maskers to provide clear examples of target-present trials. Participants were told that the target voice would not always be so apparent on the real trials and that they would have to rely on their best judgment. The instructions also stressed that this was a detection task, and that comprehension of the target voice was not required. Participants were not otherwise instructed on how to determine that the target voice was present; they were free to base their judgments on any listening strategy, which may have included listening for changes in the intensity or pattern of the speech sounds.

Participants then completed 12 practice trials. Each 3000-ms trial presented one of the 960 available maskers in either the F-F or F-RF condition. The masker onset was followed 500-1500 ms later (interval randomly chosen in ms) by either the presentation of a target or no target (SNR_{NULL}). A white fixation cross appeared in the center of the screen 250 ms before the masker onset and remained until 250 ms after the masker offset. A response prompt then followed on the screen, asking participants to press a button indicating whether or not they had detected the presence of a target on the trial. The practice consisted of 4 target-present trials (high SNRs for easy detection) and 2 target-absent trials in each spatial condition presented in random order. Participants were instructed to fixate on the white cross for the duration of each trial, and to make their response when prompted. After completing the practice and exhibiting sufficient understanding of the task, participants moved on to the screening.

Initial screening trials employed an adaptive up-down procedure to estimate the SNRs with a .50 probability of a "yes" response in each spatial condition (Levitt, 1971). F-F and F-RF trials were randomly mixed but separately tracked. On each trial, the target

word was randomly selected from the list of 80 screening words. The first SNR in an adaptive track was chosen at random, beginning with a 16-dB step size that was halved after each reversal until a step size of 2 dB was reached (Freyman et al., 2008) and then maintained for 16 more reversals. F-F and F-RF masking thresholds were then separately calculated by taking the mean SNR of the peaks and valleys across the last 16 reversals in their respective adaptive tracks. For each participant, the highest peak across the previous 16 reversals in the F-F track provided a candidate for SNR_{HIGH}, and the lowest valley provided a candidate for SNR_{SRM}. In the F-RF track, the lowest valley provided a candidate for SNR_{LOW}.

Using these results as a guide, candidate SNRs were chosen for a follow-up focused screening block, consisting of 96 trials presented in random order [8 trials x 6 SNRs x 2 spatial conditions]. Across the 80 target-present trials, each target word from the list of 80 was presented exactly once in random order. Results were assessed to determine whether the criteria for the experimental SNRs (Table 1) had been met. SNRs could be adjusted and additional blocks conducted until either the experimental SNRs could be confidently chosen or the session was concluded without moving onto the experimental section. One or two focused blocks were required to choose SNRs for all participants, with the exception of one person who received three blocks. Table 1 presents the means and *SD*s of the experimental SNRs that were chosen for the participants included in analysis.

1.2.5.2 Experimental Section

The experimental trials were identical to the screening trials in structure. EEG was recorded while participants completed 960 trials (80 trials x 6 SNRs x 2 spatial

conditions) across 10 blocks. Each block consisted of 96 trials (8 trials x 6 SNRs x 2 spatial conditions) presented in random order. Across the 960 trials, the 960 available maskers were each presented exactly once. Target words were selected from the list of 80 experimental words such that each was presented exactly once in random order within each block, and exactly once at each SNR in each spatial condition across blocks, allowing for responses to the same set of stimuli to be compared across spatial conditions at any SNR.

1.2.6 EEG Recording and Processing

Electrical Geodesics, Inc. hardware (HydroCel Geodesic Sensor Nets) and software (Net Station) were used for EEG recording and analysis. EEG was continuously recorded from 128 channels (vertex reference, 01-100 Hz bandwidth, 250 Hz sampling rate) while impedances were maintained below the recommended limit of 50 kΩ. A 60-Hz notch filter was applied to the data offline to attenuate any electrical noise acquired with the data. The EEG was then segmented into 600-ms epochs beginning 100-ms prior to the acoustic onset of each target, defined as the onset of the first appreciable burst of energy at the beginning of each word that was identified by visually inspecting the audio waveform. This onset was chosen to control for variability in the duration of the initial consonant across the targets, resulting in EEG segments that were consistently time-locked to the point in the target most relevant for detection and for eliciting a strong ERP response. Following segmentation, epochs containing data that exceeded amplitude thresholds individually set for each participant for indicating eye blinks, eye movements, or other motor movements, were excluded from analysis. The remaining artifact-free epochs were averaged together to create ERPs elicited by targets at each SNR in each

spatial condition. The ERPs were re-referenced to the average amplitude of the two mastoid channels and baseline corrected to the 100-ms interval preceding the acoustic onset of the targets.

1.2.7 Statistical Analysis

1.2.7.1 Behavior

Assessment of behavioral performance was based on the proportion of "yes" responses at each SNR in each spatial condition. In addition, the d' statistic [z(hit rate) – z(false alarm rate)] was calculated at each SNR to obtain a measure of detection accuracy independent of response bias (Green & Swets, 1966; Macmillan & Creelman, 2005). The false alarm rate at SNR_{NULL} in the respective spatial condition was used in all d' calculations. A log-linear transformation was applied to the data when calculating d' to correct for extreme response rates (Hautus, 1995).

1.2.7.2 ERPs

Visual inspection of the grand average ERP waveforms guided the approach to analysis. To assess the observed differences and their scalp distributions, a subset of 81 electrodes evenly distributed in a 3 [Anterior (A), Central (C) Posterior (P)] x 3 [Left (L), Medial (M), Right (R)] grid across the scalp was chosen for analysis. For each participant, the mean amplitude of ERPs was measured across four time windows covering the regions of what are typically the first positive-going peak (P1: 20-60 ms), first negative-going peak (N1: 130-180 ms), second positive-going peak (P2: 230-330 ms), and third positive-going peak (P3: 330-500 ms). In addition, N1-P2 latency differences found in prior studies of virtual separation for targets presented above the

informational masking threshold (Zhang et al., 2014, 2019), and visually indicated in the grand average waveforms at SNR_{HIGH}, motivated measurement of the 50% fractional negative peak latencies of ERPs 70-200 ms (N1) and positive peak latencies 200-330 ms (P2) to determine whether the onset latencies of the N1 and P2 deflections at SNR_{HIGH} differed across spatial conditions (Luck, 2014). Latencies were computed for each participant using a jackknifing procedure to reduce noise (Smulders, 2010). All measurements were collapsed across the 9 electrodes in each cell of the 3x3 grid and entered into a whole-scalp repeated-measures analysis of variance (ANOVA) model with Spatial Condition (F-F, F-RF), Anterior-Poster electrode position (A, C, P), and Left-Right electrode position (L, M, R) entered as three independent factors. In cases where a main effect of Spatial Condition was not found across the whole scalp, any significant Spatial Condition x electrode position interaction motivated follow-up analyses on subsets of electrode positions to examine whether localized effects were present. While the uncorrected degrees of freedom are reported, the Greenhouse-Geisser correction was applied to all p-values. Since the ERPs obtained at and ±2 dB from SNR_{SRM} were all found to exhibit the same pattern of effects, the reporting of results has been simplified by presenting only the ERP analyses at SNR_{SRM}, where the behavioral responses for every participant satisfied the criteria as originally intended (Table 1).

1.3 Results

1.3.1 Behavior

Masking thresholds obtained from the preliminary screening showed that all twenty participants exhibited a large release from masking (range: 14.75 – 29.00 dB SNR

of release) in the F-RF condition ($M_{threshold}$ = -19.97, SD = 4.80 dB SNR) compared to the F-F condition ($M_{threshold}$ = +2.14, SD = 1.97 dB SNR). Figure 2 shows the mean performance (proportion of "yes" responses and d' values embedded in the bars) in each spatial condition at the chosen SNRs that were presented in the experiment (Table 1). As expected from the criteria for selecting SNRs, performance was poor for targets presented below SNR_{HIGH} in the F-F condition, and good for targets above SNR_{LOW} in the F-RF condition. At SNR_{HIGH}, performance was good in both spatial conditions, but still better for all participants in the F-RF condition compared to the F-F condition. At SNR_{LOW}, although the proportion of "yes" responses did not statistically differ between spatial conditions (p = .51), d' was higher in the F-RF condition compared to the F-F condition [$t(19) = 4.37$, $p < .001$, $d = .98$], driven by a reduction in the false alarm rate (SNR_{NULL}) in the F-RF condition [$t(19) = -4.18$, $p = .001$, $d = -.94$]. Importantly, the largest benefit of spatial separation was exhibited at and around SNR_{SRM}.

1.3.2 ERPs

1.3.2.1 SNR_{LOW}

Figure 3 shows the grand average ERPs obtained at SNR_{LOW} in the F-F and F-RF conditions. The grand average waveforms were similarly nondescript and low in amplitude in both spatial conditions. Whole-scalp repeated-measures ANOVAs found no differences between the spatial conditions in mean amplitudes over the P1 (20-60 ms), P2 (230-330 ms), and P3 (330-500 ms) time windows (Spatial Condition main effect and interactions with electrode positions: ps ≥ .09). In the N1 (130-180 ms) time window, there was a three-way interaction between Spatial Condition and the two electrode

position factors [$F(4, 76) = 2.86$, $p = .04$, $\eta_p^2 = .13$], but no main effect of Spatial Condition across the scalp ($p = .85$), or at any of the subsets of electrode position ($ps \geq .19$).

1.3.2.2 SNR_{SRM}

Figure 4 shows the grand-average ERPs time-locked to targets presented at SNR_{SRM} in the F-F and F-RF conditions. Here, a striking difference was observed between the spatial conditions. While effects of target presentation were not apparent in the F-F condition, identical targets presented in the F-RF condition elicited prominent, broadly distributed N1-P2 waveforms characteristic of cortical auditory evoked potentials (N1: ~165 ms, P2: ~270 ms) and a sustained positivity across the P3 time window that was largest over the right hemisphere. In the P1 time window, there were no differences between the F-F and F-RF conditions (Spatial Condition main effect and interactions with electrode position factors: $ps \geq .18$). Targets elicited an N1 in the F-RF condition compared to the F-F condition [Spatial Condition main effect: $F(1, 19) = 23.13$, $p < .001$, $\eta_p^2 = .55$], with the largest differences observed at central and medial electrode positions [Spatial Condition x Anterior-Posterior: $F(2, 38) = 5.23$, $p < .02$, $\eta_p^2 = .22$; Spatial Condition x Anterior-Posterior x Left-Right: $F(4, 76) = 4.30$, $p = .008$, $\eta_p^2 = .18$]. Targets also elicited a P2 in the F-RF condition compared to the F-F condition [Spatial Condition main effect: $F(1, 19) = 7.97$, $p = .01$, $\eta_p^2 = .30$], with the largest differences again observed at central and medial electrode positions [Spatial Condition x Anterior-Posterior: $F(2, 38) = 4.46$, $p = .03$, $\eta_p^2 = .19$; Spatial Condition x Anterior-Posterior x Left-Right: $F(4, 76) = 2.98$, $p = .04$, $\eta_p^2 = .14$]. Although the P3 in response to targets in

the F-RF condition compared to the F-F condition was not evident across the entire scalp (p = .09), targets in the F-RF condition did elicit a P3 over the right hemisphere [Spatial Condition x Left-Right interaction: $F(2, 38) = 14.57$, $p < .001$, $\eta_p^2 = .43$; Spatial Condition x Left-Right interaction with midline sites excluded: $F(1, 19) = 23.96$, $p < .001$, $\eta_p^2 = .56$; Spatial Condition main effect at right electrodes only: $F(1, 19) = 13.83$, $p = .001$, $\eta_p^2 = .42$].

1.3.2.3 SNR_{HIGH}

Figure 5 shows the grand-average ERPs elicited by targets at SNR_{HIGH}. Broadly distributed cortical auditory evoked potentials were visible in both the F-F (N1: ~140 ms, P2: ~280 ms) and F-RF (N1: ~130 ms, P2: ~270 ms) conditions. No differences in mean amplitude between spatial conditions were observed in the P1 time window (ps ≥ .14). However, ERPs elicited by targets in the F-RF condition compared to the F-F condition were more negative in the N1 window [$F(1, 19) = 5.99$, $p = .02$, $\eta_p^2 = .24$] and more positive in the P2 window [$F(1, 19) = 4.37$, $p = .05$, $\eta_p^2 = .19$]. In the P3 time window, there were interactions between Spatial Condition and electrode position factors [Spatial Condition x Left-Right: $F(2, 38) = 12.63$, $p < .001$, $\eta_p^2 = .40$; Spatial Condition x Anterior-Posterior x Left-Right: $F(4, 76) = 3.29$, $p < .03$, $\eta_p^2 = .15$], but no main effect of Spatial Condition across the scalp (p = . 73), or at any of the subsets of electrode position (p's ≥ . 08). The 50% fractional peak latencies were shorter for the N1 [$F(1, 19) = 6.21$, $p = .02$, $\eta_p^2 = .25$] and P2 [$F(1, 19) = 9.73$, $p = .006$, $\eta_p^2 = .34$] in the F-RF condition compared to the F-F condition.

1.4 Discussion

The present study captured dramatic changes in neural processing underlying spatial release from informational masking. The use of noise-vocoding to minimize non-spatial cues and a behavioral screening for selecting the experimental stimuli allowed for the comparison of ERPs elicited by the same targets when they were masked in the spatially co-located condition and unmasked in the virtually separated condition. Results suggest that within noisy, complex environments, spatial separation alone can provide a powerful cue for reducing the informational masking caused by target/masker confusion and improving target representations beginning early in perceptual processing.

1.4.1 Behavior

Thresholds obtained from the adaptive procedure in the screening section showed that all participants exhibited a large release from informational masking across spatial conditions consistent with prior virtual separation studies with noise-vocoded speech (Freyman et al., 2008; Morse-Fortier et al., 2017; Zobel et al., 2019). Importantly, thresholds in the F-F condition were consistently in the range around 0 dB SNR posited to be the ceiling for informational masking (Arbogast et al., 2005; Freyman et al., 2008). Maximal informational masking in the F-F condition suggests that the vocoded stimuli were effective at keeping beneficial non-spatial cues to a minimum and that the masking release observed in the F-RF condition was specific to the spatial cue alone.

The behavioral results from the experimental section confirmed that the screening procedure was effective at identifying the set of experimental SNRs designed to probe the ERP effects of spatial release from informational masking. Poor detection performance at SNR_{LOW} in both spatial conditions suggested that the low-intensity targets were

presented below participants' energetic masking thresholds. At SNR_{SRM}, targets were largely masked in the F-F condition and largely unmasked in the F-RF condition, providing an ideal comparison for measuring the neural processing specific to spatial release from informational masking. Performance at SNR_{HIGH} was generally good in both spatial conditions, suggesting that these high-intensity targets were consistently presented above the informational masking threshold, although enough informational masking was still present in the F-F condition to observe some amount of spatial release in the F-RF condition.

The yes/no task was useful in providing measures of both the hit rates and the false-alarm rates. Spatially separating targets and maskers in the F-RF condition improved participants' ability to recognize both when the target was present (increased hit rate) and when the target was absent (decreased false-alarm rate), a pattern previously reported under similar conditions in younger and older adults (Zobel et al., 2019). A simple hypothesis can be offered for these results based on a reduction in informational masking (i.e., a reduction in target/masker confusion): When target and masker were spatially co-located, participants could easily confuse the presence of a target as a fluctuation in the masker (decreased hit rate) and a fluctuation in the masker as the presence of a target (increased false-alarm rate). When target and masker were spatially separated, however, participants could adopt the simple but effective strategy of responding "yes" to any sound heard from the front and responding "no" otherwise, resulting in improvements to both the hit and false-alarm rates. This hypothesis is consistent with research suggesting that a spatial cue releases informational masking by facilitating the segregation of target and masker into independent auditory objects at

separate locations that can be selectively attended (Ihlefeld & Shinn-Cunningham, 2008b, 2008a), and suggests that listening strategy may also be an important factor to consider in the spatial release from informational masking.

1.4.2 ERPs

1.4.2.1 SNR_{LOW}

No systematic effects were observed in the ERPs time-locked to targets presented at SNR_{LOW} in either spatial condition. This is consistent with the behavioral data suggesting that the low-intensity target information was below the energetic masking threshold (constant across virtual separation conditions), and thus unavailable to elicit a cortical response. Importantly, similar low-amplitude non-descript waveforms were observed in both spatial conditions; no confounding ERP effects that may have been driven by the physical changes across conditions (i.e., the addition of the masker at the right in the F-RF condition) were observed. Thus, ERPs at SNR_{LOW} provide a baseline for assessing effects specific to informational masking when targets were presented above the energetic masking threshold at SNR_{SRM} and SNR_{HIGH}.

1.4.2.2 SNR_{SRM}

A striking increase in the amplitude of the cortical auditory evoked potentials elicited by targets at SNR_{SRM} was observed when targets and maskers were virtually separated compared to spatially co-located. This result was consistent with prior studies showing an increase in N1/P2 amplitude for virtually separated targets (Zhang et al., 2014, 2019). However, the extent to which this effect was modulated across the spatial conditions in the present study has not been previously reported. The prior ERP studies

of virtual separation presented targets above the informational masking threshold where they elicited clear cortical auditory evoked potentials in both spatial conditions (Zhang et al., 2014, 2019). In contrast, targets at SNR_{SRM} in the present study were masked in the F-F condition and unmasked in the F-RF condition, allowing for the ERP effects of informational masking and its spatial release to be better isolated. Indeed, when SNR_{SRM} targets were masked in the F-F condition, despite being presented well above the energetic masking threshold (M = 17.57 dB SNR above F-RF threshold, SD_{Diff} = 5.0 dB SNR), ERPs exhibited little, if any, effects of target presentation. When the same set of targets were released from masking in the F-RF condition, cortical auditory-evoked potentials were apparent. Taken together, these results show that informational masking can severely limit target representation, and that the spatial cue alone can greatly facilitate target representation, and that both of these aspects of spatial release from informational masking can occur early in perceptual processing, at least as early as the N1. These effects are also broadly consistent with evidence suggesting that spatial separation reduces informational masking by facilitating bottom-up processes involved in auditory object formation and top-down selective attention (Ihlefeld & Shinn-Cunningham, 2008b, 2008a; Zhang et al., 2014, 2019). Auditory streaming, object formation and segregation, and selective attention have all been shown to influence ERP amplitude within the N1-P2 time window (Alain et al., 2002; Dyson et al., 2005; Hautus & Johnson, 2005; Hillyard et al., 1973; Snyder et al., 2006; Woldorff & Hillyard, 1991; Zobel et al., 2015). Furthermore, prior ERP studies of virtual separation found that N1/P2 effects of spatial separation were enhanced by attention but still present (particularly the N1 effect) when attention was directed away from the targets (Zhang et

al., 2019). In the present study, however, cortical auditory evoked potentials were only apparent in the F-RF condition, and the relative contributions of bottom-up and top-down processing in accounting for this effect of unmasking will require further investigation.

It is possible that processes responsible for generating the cortical auditory-evoked potentials themselves were modulated across spatial conditions. The N1-P2 response reflects neural activity in auditory cortex that is generally described as obligatory or exogenously driven by a physical event in the auditory environment, such as the onset of a target word (Kraus & Nicol, 2008; Lightfoot, 2016; Picton, 2013). In the present study, however, identical targets were presented in both spatial conditions while holding sensory effects (i.e., energetic masking) constant. The fact that such a large modulation of N1-P2 amplitude was driven solely by the release from informational masking, therefore, suggests that the cortical auditory evoked potential itself may depend more heavily upon how the auditory scene is perceptually organized rather than physically organized. If so, the present results would be consistent with the proposition that energetic and informational masking both describe interference produced when target and masker elicit overlapping patterns of neural activity, with the main difference being the stage of processing (e.g., sensory vs. perceptual) at which these patterns occur (Durlach et al., 2003; Zhang et al., 2014). As such, the present ERP results may imply that the neural populations responsible for the perceptual representations of targets and maskers overlap more extensively when targets and maskers are spatially co-located compared to spatially separated. Further research will be required to test this hypothesis.

In addition to the early effects discussed above, a right-lateralized late positivity was elicited by targets that were spatially separated compared to co-located with maskers.

The timing of this positive deflection is consistent with the P300, a robust component associated with the cognitive processing of a target stimulus (Polich, 2007). P300 effects were not elicited in prior ERP studies of virtual separation, because the stimulus used to obtain the ERPs differed from the stimulus that participants were asked to detect (Zhang et al., 2014, 2019). The ERPs in the present study reflect the processing of the task-relevant targets themselves, and the observed P300 suggests that spatial separation allowed listeners to better attend to and classify the target sound according to the task (Polich, 2007). It is unclear why the P300 was right-lateralized under the present conditions. Studies have shown cortical auditory evoked potentials and P300 effects that are larger over the hemisphere contralateral to the stimulus location (Gilmore et al., 2009; Wolpaw & Penry, 1977). Although the target was always presented from the front in the present study, the addition of the masker energy from the right in the F-RF condition may have resulted in a better ear for listening on the left and a larger contralateral P300. The effect may not have been strong enough to observe in the cortical auditory evoked potentials, but the P300, being a larger and later component, may have been more sensitive to this distinction. Further research will be needed to determine the precise relationship between the location of sounds and the distribution of early and late ERP responses under the present conditions.

1.4.2.3 SNR_{HIGH}

Targets presented at SNR_{HIGH} elicited visible cortical auditory evoked potentials in both spatial conditions, consistent with the behavioral data showing that SNR_{HIGH} was above the informational masking threshold. The N1/P2 waveforms at SNR_{HIGH} were larger and began earlier in time in the F-RF condition compared to the F-F condition—

also consistent with the behavioral data showing some benefit of spatial separation—while a P300 effect was not evident. Taken together, these results suggest that although participants could perform the task well and detect the majority of targets in both spatial conditions, the perceptual representation of targets was still consistently better when targets were spatially separated compared to co-located with maskers. The results directly replicate the effects reported in the prior ERP studies of virtual separation (Zhang et al., 2014, 2019). In the prior studies, where all targets were presented above the informational masking threshold, the researchers separately interpreted the N1/P2 amplitude and latency effects, positing that amplitude may be related to target/masker segregation or object grouping while latency may be related to selective attention or listening effort (Zhang et al., 2014, 2019). Although object formation and selective attention may contribute to these early effects, the strong N1/P2 modulation observed at SNR_{SRM} in the present study allows the amplitude and latency effects at SNR_{HIGH} to be explained by a single index of spatial release from informational masking. Considering that each target onset (or the /ba/ syllable onset used in the prior studies) contains an energy burst that ramps up over a given period of time, target intensity should reach an SNR that exceeds masking threshold sooner in the F-RF condition compared to the F-F condition. If the cortical auditory evoked potentials are only driven by the unmasked portion of the target sound, as observed at SNR_{SRM}, then targets at SNR_{HIGH} should elicit cortical auditory evoked potentials earlier in time when the targets are spatially separated compared to spatially co-located with maskers. In short, these results suggest that spatial release from informational masking can be characterized by an improvement in the early perceptual representation of a target sound over a wider range of intensity that will

typically extend to portions of the target onset that occur earlier in time. These early perceptual benefits of spatial separation are evident for targets presented well above the threshold of detection, and are likely to support higher-level processes involved in the recognition and comprehension of speech under challenging listening conditions.

In conclusion, the present study offered detailed insight into the processing of identical speech sounds under conditions of informational masking and its spatial release. ERPs revealed dramatic perceptual benefits when a spatial cue alone was introduced to resolve target/masker confusion and cognitive benefits associated with classifying task-relevant sounds. These findings suggest that within challenging listening environments, spatial separation can help listeners solve the cocktail party problem by providing a powerful cue for reducing perceptual confusion and improving the early perceptual representation of relevant speech that supports later stages of cognitive processing. The striking effects observed when informational masking was released also suggest that the cortical auditory evoked potential itself is highly sensitive to the perceptual organization of stimulus information. Together, these results establish ERP indices of spatial release from informational masking that can be used to further investigate its underlying mechanisms. Study 2 was designed to measure how these indices differ when listeners are attending to and away from the sounds to assess the relative contributions of preattentive and attentive processing in the spatial release from informational masking.

CHAPTER 2

STUDY 2: THE EFFECTS OF ATTENTION ON THE SPATIAL RELEASE FROM INFORMATIONAL MASKING

2.1 Introduction

Results from Study 1 showed that under challenging listening conditions, spatial separation between competing speech streams can reduce target/masker confusion and benefit listeners at early perceptual (N1/P2 amplitude and latency effects) and later cognitive (P300 effect) stages of processing. These ERP indices of spatial release from informational masking are useful for further exploring the underlying mechanisms that contribute to these benefits. Research suggests spatial release from informational masking reflects the bottom-up processing of target and masker into separate auditory objects, and top-down selective attention to the target, but the relative contributions of bottom-up and top-down mechanisms, the stages of processing in which they operate, and the extent to which they are interdependent are not fully understood. How much of the effects observed in Study 1 depended on the fact that the participants were always directing their attention to the sounds? Experiment 2 was designed to examine the role of attention by measuring how the ERP indices of spatial release from information change when listeners direct their attention away from the auditory modality. Research on the underlying mechanisms that contribute to spatial release from informational masking may offer avenues for future investigation of listeners who struggle with informational masking and experience difficulties processing speech within complex listening environments, including children and older adults, and individuals with hearing loss

(Allen & Wightman, 1995; Feng et al., 2018; Huang et al., 2010; Wightman et al., 2003; Wightman & Kistler, 2005; M. Wu, Li, Hong, et al., 2012; Zobel et al., 2019).

A large body of research on auditory scene analysis asserts that the perceptual organization of an auditory scene involves bottom-up processes that automatically group sounds into distinct auditory objects based on their related features (Bregman, 1990). This can be demonstrated in the lab, for example, when listeners are presented with a sequence of alternating low- and high-frequency pure tones and automatically perceive the sequence to split into two distinct auditory streams, one consisting of repeating low-frequency tones, and the other consisting of repeating high-frequency tones. Similar auditory object grouping can be demonstrated for simultaneously presented sounds. For example, the concurrent presentation of a set of harmonically related pure tones will be automatically perceived as a single unified complex sound, and mistuning one of the harmonics will cause it to perceptually split apart from the complex sound as a separate auditory object (Alain et al., 2001, 2002; Alain & Izenberg, 2003; Bregman, 1990; Dyson et al., 2005). In addition to frequency, many distinguishing cues have been shown to facilitate auditory object grouping, such as timing, intensity, timbre, and spatial location (Bregman, 1990). Behavioral evidence suggests that this auditory object grouping involves preattentive processes. Listeners often cannot control how the auditory input is perceptually parsed into objects, and once objects are formed, listeners may not be able to report upon certain aspects of their individual components, such as the relative positions of tones across streams, or the individual harmonics within a complex sound (Bregman, 1990). ERPs have provided additional insight into the bottom-up mechanisms of object grouping. Presenting a complex periodic sound containing one mistuned harmonic that is

perceived as a separate object elicits ERPs that are more negative 150-250 ms after sound onset compared to presenting the same complex sound with all harmonics tuned and perceived as a single unified object (Alain et al., 2001, 2002; Alain & Izenberg, 2003; Dyson et al., 2005). A similar negativity is observed when identical broadband sounds are presented to both ears with a narrow band containing an interaural time difference that results in the perception of two auditory objects at separate locations, compared to the same broadband sounds presented with no interaural time difference, resulting in the perception of a single fused object at one location (Hautus & Johnson, 2005). This negativity is thought to index the general perception of two auditory objects compared to one, and has been dubbed the object related negativity (ORN) (Alain et al., 2001). Studies have shown that under such conditions of object grouping, the ORN is elicited regardless of whether listeners are attending to the sounds or attending to a book or silent movie (Alain et al., 2001, 2002), the opposite ear in a dichotic listening task (Alain & Izenberg, 2003) or simple visual tasks (Dyson et al., 2005), suggesting that the underlying mechanisms can operate in the absence of attention. Given this evidence, it follows that under challenging listening conditions, the spatial separation between target and masker provides a distinguishing cue that facilitates the bottom-up grouping of target and masker sounds into separate auditory objects, thus reducing their perceptual confusion and releasing informational masking. The dramatic effects of unmasking in the F-RF condition that were observed in Study 1 at early perceptual stages of auditory processing suggest some contribution from bottom-up processing. Bottom-up object formation, however, may not account for the entirety of a listener's benefit when target and masker are spatially separated.

Research also shows that top-down attention can play an important role in informational masking and its spatial release. A long history of research demonstrates the importance of selective attention in solving the cocktail party problem, dating back to when the problem was first proposed (Cherry, 1953). Dichotic listening studies on the cocktail party problem have been useful in describing how listeners can take advantage of a separation between target and masker to direct their attention to the target and ignore the masker, and their findings have informed the development of important models of selective attention (Broadbent, 1958; Pashler, 1999; Treisman, 1964). Research specific to informational masking suggests that target/masker confusion may arise from a lack of predictable structure within the auditory information that would allow a listener to better attend to the target and ignore the masker. For example, informational masking is sensitive to the degree of uncertainty (i.e., unpredictability) surrounding the auditory stimuli and its presentation. An increase in the trial-by-trial variability of the auditory stimuli will produce an increase in informational masking, while a reduction in variability will release informational masking (Kidd et al., 2008; Lutfi, 1993; Lutfi et al., 2013; Watson et al., 1975, 1976). Indeed, despite maximal informational masking in the F-F condition in Study 1, unpublished data obtained under similar conditions suggests that a substantial reduction in F-F masking threshold would have been observed across trials if listeners had been presented with the same exact target on each trial (R. Freyman, personal communication, 2019). Likewise, the masking release observed in Study 1 likely depended to some extent on the target location remaining constant throughout the experiment; research suggests that spatial release from masking would have been reduced if target location had been stochastically varied from trial to trial (Brungart & Simpson,

2007; Ericson et al., 2004; Kidd, Arbogast, et al., 2005). The benefit of greater predictability observed across trials is likely to arise from the fact that listeners come to learn about certain features of the target to which they can direct their attention in anticipation of its presentation (Watson et al., 1976). Research has also shown that informational masking is sensitive to the degree of perceptual similarity between target and masker (e.g., higher informational masking for same-sex compared to opposite-sex target and masker) (Brungart, 2005; Brungart et al., 2001; Festen & Plomp, 1990). Lutfi et al. (2013) relates these effects of uncertainty and perceptual similarity under a common measure of information divergence, arguing that informational masking is not specific to the acoustic properties or auditory context, but generally arises under any condition in which there is a lack of discernable statistical distinctions between the target and masker information. This explains why listeners can take advantage of a variety of different cues as long as they predictably distinguish targets and maskers (Başkent & Gaudrain, 2016; Bradlow & Alexander, 2007; Brungart, 2005; Brungart et al., 2001; Cherry, 1953; Culling & Summerfield, 1995; Darwin et al., 2003; Darwin & Hukin, 2000; El Boghdady et al., 2019; Ericson et al., 2004; Freyman et al., 2001, 2004, 2005; Kidd, Arbogast, et al., 2005; Kidd et al., 1994; Mattys et al., 2012; Vestergaard et al., 2009; Watson et al., 1975). Among the many ways listeners may benefit from cues that provide statistical distinctions within the auditory information, such cues equip listeners with reliable knowledge about the auditory scene that allows them to effectively direct their attention in anticipation of relevant information. Indeed, research has shown greater release from informational masking when listeners are explicitly primed with a priori knowledge that would facilitate selective attention, such as knowledge about the acoustic features and

content of the target and masker (Feng et al., 2018; Freyman et al., 2004; Huang et al., 2010; Newman & Evers, 2007; Richards et al., 2004; Richards & Neff, 2003; Singh et al., 2008; M. Wu, Li, Gao, et al., 2012; M. Wu, Li, Hong, et al., 2012; Yang et al., 2007; Yonan & Sommers, 2000), and knowledge about the spatial location of the target (Ericson et al., 2004; Kidd, Arbogast, et al., 2005). Top-down attentional effects on informational masking are also indicated in studies showing a listening advantage for trained vs. untrained musicians under conditions of informational masking (Morse-Fortier et al., 2017; Oxenham et al., 2003), ERP research showing greater attentional modulation of the cortical auditory evoked potentials elicited by targets under conditions of informational compared to energetic masking (Zhang et al., 2016), and developmental research showing that children are more susceptible to informational masking than young adults (Allen & Wightman, 1995; Wightman et al., 2003; Wightman & Kistler, 2005), and that older adults exhibit declines in the ability to use a priori knowledge (Feng et al., 2018; Huang et al., 2010; M. Wu, Li, Hong, et al., 2012) and spatial separation (Zobel et al., 2019) to release informational masking. In addition, results from Study 1 may indicate an effect of selective attention on spatial release from informational masking. Spatial separation improved not only participants' ability to recognize when a target was present on a trial, but also when a target was absent, suggesting that listeners took advantage of the spatial cue to better attend to the front location and minimize interference from the masker sounds.

Together, this research supports a hypothesis of the bottom-up and top-down mechanisms that contribute to spatial release from informational masking: Within challenging listening environments, a spatial cue creates statistical distinctions within the

auditory information that facilitate the grouping of target and masker sounds into separate objects at separate locations and provide predictable structure for listeners to better direct their attention to the target and ignore the masker. Ihlefeld & Shinn-Cunningham (2008a) posited a similar model of bottom-up and top-down processing in their behavioral research on spatial release from informational masking. Participants were asked to report key words from target messages presented with masking speech. The speech stimuli were sine-wave vocoded to minimize spectral overlap between target and masker and ensure that masking was predominantly informational rather than energetic. Targets and maskers were varied in their spatial location and timbre, while participants were asked in separate conditions to attend to the target location (target timbre not known a priori), timbre (target location not known a priori), or location and timbre (target location and timbre both known a priori), and to report the target content. Results showed that performance improved with increasing spatial separation between target and masker under the two conditions in which attention was cued to location, but remained flat under the condition in which attention was cued only to timbre, suggesting that spatial separation facilitated spatially selective attention to the target. However, increasing spatial separation reduced errors indicative of perceptually confusing the target and masker streams in all three attention conditions, suggesting that the spatial cue also facilitated the grouping of target and masker into separate objects, regardless of how participants directed their attention. These results were consistent with those from a separate study examining differential effects of energetic and informational masking under similar stimulus conditions (Ihlefeld & Shinn-Cunningham, 2008b). In both studies, however, confounding alternative explanations limited the authors from drawing

definitive conclusions about the extent to which the spatial cue facilitated processing independently of attention. These studies provide a useful framework for understanding the mechanisms of spatial release from informational masking that help listeners solve the cocktail party problem, but they also highlight the common challenges of assessing the relative contributions of preattentive and attentive processing through behavioral measures alone (Pashler, 1999). Disentangling bottom-up and top-down effects is especially challenging given that attention may modulate bottom-up processing under complex listening conditions. For example, although auditory streaming occurs automatically when strong cues are present, weaker or more ambiguous cues may allow listeners to control how sounds are grouped into separate streams and to switch between different percepts (Bregman, 1990). Neuropsychological research on patients with unilateral neglect also suggests that attention can play a crucial role in auditory object formation (Carlyon et al., 2001). Therefore, to assess the extent to which bottom-up and top-down processes contribute to spatial release from informational masking, a precise measure of auditory processing is needed that can be obtained in the presence and absence of auditory attention without requiring a behavioral response.

ERPs can provide such a measure, and have been used for decades to explore the role of attention in auditory processing. ERPs have revealed extensive effects of attention at early perceptual and later cognitive stages of auditory processing. For example, attended sounds tend to elicit larger auditory evoked potentials, with a sustained negative difference wave (Nd) that can extend beyond the N1 time window, and a P300 to targets compared to unattended sounds (Luck & Kappenman, 2011). ERPs have also been useful for revealing effects of attention on auditory object formation under more

complex listening conditions. Research on the precedence effect, for example, has shown that the ORN observed for attended sounds at and around echo threshold (SOA at which lead and lag sounds are heard as separate objects at separate locations), was not observed for unattended sounds, suggesting that attention is necessary for object formation in the precedence effect. Several factors may explain why attentional modulation of the ORN was observed in the precedence effect and not in prior studies of harmonic grouping and dichotic pitch, including that a stronger attention manipulation was used to direct attention away from the auditory modality (difficult two-back visual task), that grouping cues may be weaker or more ambiguous near echo threshold, and that the precedence effect may require more complex processing in order to group sounds into objects from different locations across relatively long delays (Zobel et al., 2015). This study shows that ERPs are well-suited for examining the role of attention at precise stages of auditory processing under complex listening conditions involving sound localization and object grouping.

To examine the role of attention in the spatial release from informational masking, the present study used ERPs to measure responses to targets presented within the virtual separation paradigm while in separate conditions participants attended to the sounds and attended away from the auditory modality. The prior ERP studies of virtual separation included attention manipulations (Zhang et al., 2014, 2019). Zhang et al. (2014) found that under conditions of informational masking, larger auditory evoked potentials were elicited in the spatially separated condition regardless of whether listeners attended to the sounds or attended to a silent movie, but spatial effects on N1 and P2 latency were only present when listeners attended to the sounds. However, statistical

analyses of the two-way interactions between the spatial and attention conditions were not reported, and the extent to which attention modulated early perceptual effects of spatial release from informational masking is unclear. Zhang et al. (2019) examined the two-way interactions between spatial separation and spatially selective attention. Results showed that attention modulated spatial effects on P2 amplitude but not N1 amplitude, while no modulation of latency effects was found in either time window. These results are generally consistent with the two-stage model described above (Ihlefeld & Shinn-Cunningham, 2008b, 2008a) in which spatial separation elicits an automatic grouping of target and masker sounds into separate objects at separate locations (N1 time window) that can then be selectively attended by the listener (P2 time window). However, it is difficult to assess the extent to which attentional modulation of early effects are consistent across these studies, especially since Zhang et al.'s (2014) results indicate possible interactions between spatial separation and attention beginning within the N1 time window. Most importantly, the stimuli in both of these studies were presented well above the informational masking threshold in all conditions. The role of attention in the spatial release from informational masking is impossible to determine when a strong non-spatial cue (SNR) is already present for separating the target from the masker. In the present study, the same stimuli from Study 1 were used to ensure that non-spatial cues were minimized. The same screening procedure was used prior to beginning the experiment to choose for each participant an SNR at which targets would be masked in the F-F condition and unmasked in the F-RF condition (SNR_{SRM}), as well as an SNR above masking threshold in both spatial conditions (SNR_{HIGH}). In one condition of the present study, participants engaged in the target detection task described in Study 1. In a

separate condition, attention was directed away from the auditory modality by having participants engage in a challenging two-back visual task similar to one used in prior research on attention in the precedence effect (Zobel et al., 2015). Comparison of the spatial effects on ERPs time-locked to the onset of targets presented within each condition provided detailed measures of the preattentive and attentive mechanisms that contribute to the spatial release from informational masking.

2.2 Methods

Study 2 similarly comprised two sections: 1) a screening section similar to Study 1, consisting of a hearing assessment and the collection of behavioral data to measure masking thresholds and determine the experimental SNRs and 2) an experimental section in which behavioral and EEG data were collected while in separate blocks of trials, participants attended to the auditory target detection task and attended to a difficult two-back visual task. All participants except for one (described below under "screening section") completed the study in a single ~3-hr session.

2.2.1 Participants

Eighteen right-handed English-speaking adults (8 female) aged 20-35 years (M = 26.39, SD = 4.12 years) reporting no known hearing, visual, or neurological problems and no use of psychoactive medication at the time of the study, contributed the data for analysis. Data from an additional nine participants were excluded because at least one of their pure-tone audiometric hearing thresholds exceeded 20 dB HL at either 1, 2, 4, or 8 kHz (n =1), their responses in the screening section did not exhibit spatial release from masking (n = 2) or allow the experimental SNRs to be reliably determined (n = 1), they

fell asleep (n = 1), their EEG contained excessive noise related to ocular and motor movements, muscle tension, low-frequency oscillations, and/or electrode bridging (n = 3), or the equipment malfunctioned (n = 1). All participants provided informed consent prior to beginning the study and were compensated at a rate of $10/hr.

2.2.2 Stimuli

2.2.2.1 Auditory Stimuli

Auditory targets were the same single-syllable noise-vocoded words used in Study 1. Again, the 80 screening words were used in the screening section, and the 80 experimental words were used in the experimental section. Maskers were two-talker noise-vocoded babble created from the same recordings of nonsense sentences used in Study 1. This time, the recordings were each evenly divided into 1280 individual 2000-ms segments that were then randomly paired and vocoded as described in Study 1. This resulted in 1280 different 2000-ms vocoded two-talker maskers (note that masker length was 500 ms shorter than in Study 1 to accommodate an increase in experimental trials).

The same procedure described in Study 1 was used to create the auditory target and masker stereo WAV files (44.1 kHz sampling rate, 16-bit resolution). Each auditory target was available to be presented in 1-dB steps from -40 dB to +10 dB relative to the masker RMS amplitude. Each masker was available to be presented in the F-F and F-RF formats. Thus, again, the auditory target intensity was varied while masker intensity was held constant throughout the experiment, and the reported SNRs in both spatial conditions were calculated as the RMS amplitude of the auditory target relative to the masker presented from the front in all conditions.

2.2.2.2 Visual stimuli

The visual stimuli consisted of 13 different letters (a, b, d, e, f, g, h, j, m, n, q, r, and t), each in a white font against a black background.

2.2.3 Study Setup

The experiment was conducted in the same room with the same equipment and setup used in Study 1. Participants were seated 1.5 m away from the two loudspeakers that were positioned 0° and 55° right of midline, respectively, and each set to present the maskers at 70 dBA SPL at the listening position. The visual stimuli were displayed in the center of the computer monitor positioned directly beneath the front loudspeaker such that each letter subtended < 1° of visual angle at the viewing position.

2.2.4 Procedure

2.2.4.1 Screening section

The same screening procedure described in Study 1 was used to determine F-F and F-RF masking thresholds and the experimental SNRs for each participant in Study 2. Since the masker length had been shortened to 2000 ms, the individual trials were 2500 ms in duration with a 400-1400-ms SOA between masker and auditory target. In all other respects, screening trials were identical to those described in Study 1. Following the hearing assessment, and instructions on the auditory task, participants completed the adaptive screening procedure to estimate their masking thresholds and subsequent focused screening blocks of trials to determine their experimental SNRs. To accommodate the larger number of experimental trials required for Study 2, only the three SNRs most essential to the results of Study 1 were chosen to be presented in Study

2: $SNR_{NULL,}$ SNR_{SRM}, and SNR_{HIGH}. No more than three focused blocks were required to choose SNRs for all participants. Table 1 presents the means and *SD*s of the experimental SNRs that were chosen for the participants included in analysis.

2.2.4.1.1 Noteworthy participants

Initially, participants were not asked to move onto the experimental section if their adaptive screening thresholds did not show a spatial release from masking. However, this requirement was dropped after excluding two participants early on in recruitment whose adaptive screening results may have been skewed by a liberal response bias (i.e., detection-independent tendency to respond "yes" in the F-F condition) (Green & Swets, 1966; Macmillan & Creelman, 2005). Subsequently, all participants were allowed to complete the experiment as long as their focused screening (which allowed hits and false alarms to be evaluated) could reliably determine their experimental SNRs. As a result, there was one participant that contributed data for analysis whose adaptive screening showed an F-F threshold (-37.50 dB SNR) that was lower than their F-RF threshold (-21.63 dB SNR). Their focused screening, however, when accounting for a liberal response bias in the F-F condition, determined their SNR_{SRM} to be -2 dB SNR, and their experimental results confirmed that they indeed exhibited a substantial spatial release from masking at SNR_{SRM} in the F-RF condition (d' = 3.27) compared to the F-F condition (d' = 0.46). Data from this participant were excluded from analyses involving the adaptive screening results, but included in all other analyses.

Another participant who contributed data for analysis completed the experiment in two separately scheduled sessions instead of one. The first session was stopped early due to an equipment malfunction that occurred just after the participant had completed

the screening section. Upon returning to the lab for the second session, the participant again received the instructions on the auditory task and then completed two focused screening blocks before moving on to complete the experimental section. Data from this participant were included in all analyses.

2.2.4.2 Experimental section

The experimental section began with an introduction to the visual two-back task. The visual task consisted of a stream of letters presented one at a time (700-ms duration, 200-ms interstimulus interval) in random order in the center of the computer screen, with the letter case alternating on every other presentation (e.g., two lowercase letters, two uppercase letters, two lowercase letters, etc.). A visual target was defined as the presentation of a letter that matched the letter of the alphabet presented two positions back in the stream (note that the visual target and the letter two positions back were always opposite in letter case). Participants were asked to respond to a visual target by pressing one button (same button as a "yes" response on the auditory task) if the visual target was uppercase, and another button (same button as a "no" response on the auditory task) if the visual target was lowercase. To keep participants motivated and entertained, the task was presented as a game against the computer. A correct button press that occurred within a response window 150-1500 ms after the onset of a visual target would earn the participant a point. An incorrect button press, or no response to a visual target (miss), or any response that fell outside of the response window (false alarm) would earn the computer a point. The length of the response window was not explicitly stated in the instructions. Participants were simply encouraged to respond as quickly as they could and were instructed in general terms that a quick and correct response would earn them a

point while an incorrect response, a missed response, a slow response, or responding when a visual target had not been shown would earn the computer a point. They were also told that maintaining focus on the task was crucial because a visual target could appear at any point in the stream and more than one visual target could appear in a row. To practice the visual task, participants were first presented with a letter stream such that within every sequence of seven or eight letters (M = 7.8 letters), two of the letters were chosen at random to be a visual target. The stream continued until the participant had achieved four correct responses. Participants then practiced competitively against the computer with a 27-letter stream containing 6 visual targets. This was followed by the presentation of a score board showing the participant's points vs. the computer's points (broken down by number of correct responses, incorrect responses, misses, and false alarms). If the participant had obtained an equal or greater number of points compared to the computer, the practice was concluded; otherwise, the scores were reset to zero and the competition was repeated until this performance criterion was met. No more than five practice rounds against the computer were required for all participants to reach this criterion.

After concluding the practice with the visual task, participants were introduced to the two types of trial blocks that would be presented in the experiment. The Attend Visual trial block consisted of concurrently presented visual and auditory streams. The visual stream consisted of letters (starting letter case randomly chosen) and visual targets as described above, and the auditory stream consisted of auditory trials presented one after the other with an intertrial interval of 1500 ms. Each auditory trial was composed of a masker followed 400-1400 later by either an auditory target or no auditory target. To

avoid consistent time-locking of the visual and auditory streams across blocks, the onset of the auditory stream was delayed 0-700 ms (interval randomly chosen in ms) relative to the onset of the visual stream at the beginning of each block. A fixation cross appeared on the screen for 600 ms followed by a black screen for 200 ms before the first letter and after the last letter of the visual stream. The Attend Auditory block was identical in structure to the Attend Visual block except that the visual stream did not contain any visual targets. For the Attend Visual block, participants were told that their task was to ignore the sounds and focus solely on responding accurately to visual targets to obtain the best score they could against the computer. For the Attend Auditory block, participants were told that their task was to listen to each auditory trial and make a response during the 1500-ms intertrial interval to indicate whether or not an auditory target had been heard. If they made a response to every auditory trial in an Attend Auditory block, they would gain 5 points against the computer; if they missed responding to one or more trials in a block, they would lose 10 points. Participants were also instructed that visual targets would not be presented during Attend Auditory blocks, and that they should simply use the visual stream as a place to fixate their eyes while listening and responding to the auditory trials. Participants then completed short practice Attend Visual and Attend Auditory blocks while the experimenter watched to make sure that they were performing the tasks correctly. The Attend Visual block consisted of five visual targets, and both blocks each consisted of five auditory trials (3 target-present, 2 target-absent) presented in random order.

Following the practice, EEG was recorded while participants completed 16 blocks of trials (8 Attend Visual, 8 Attend Auditory) presented in random order, each lasting ~5

minutes. Prior to beginning each block, the visual stream was constructed so that each of the 13 letters would appear exactly 24 times in random order across a total sequence of 312 presentations without any visual targets. In the Attend Visual condition, letters within this sequence were then switched to visual targets prior to presentation, such that the Attend Visual blocks each contained 80 visual targets distributed throughout the visual stream as previously described (i.e., two visual targets randomly placed within every sequence of 7 or 8 letters), for a total of 640 visual targets presented in the Attend Visual condition. In the Attend Auditory condition, the visual sequence was presented as initially constructed without any visual targets. Both the Attend Visual and Attend Auditory blocks each contained 80 auditory trials, consisting of 40 F-F trials (10 SNR_{NULL}, 10 SNR_{SRM}, 20 SNR_{HIGH}) and 40 F-RF trials (20 SNR_{NULL}, 10 SNR_{SRM}, 10 SNR_{HIGH}) presented in random order. Note that the trial numbers were doubled at certain SNRs in each spatial condition in order to balance the number of trials expected to produce "yes" and "no" responses within each spatial condition in the Attend Auditory blocks. That is, in the F-F condition, the number of trials at the SNR expected to elicit a "yes" response (i.e., SNR_{HIGH}) were doubled to match the combined number of trials at the SNRs expected to elicit a "no" response (i.e., SNR_{NULL} and SNR_{SRM}). Likewise, in the F-RF condition, the number of trials at the SNR expected to elicit a "no" response (i.e., SNR_{NULL}) were doubled to match the combined number trials at the SNRs expected to elicit a "yes" response (i.e., SNR_{SRM} and SNR_{HIGH}). In total, 640 auditory trials were presented in each attention condition: 320 F-F trials (80 SNR_{NULL}, 80 SNR_{SRM}, 160 SNR_{HIGH}) and 320 F-RF trials (160 SNR_{NULL}, 80 SNR_{SRM}, 80 SNR_{HIGH}). The target word that was presented on each target-present auditory trial was pseudorandomly selected

from the list of 80 experimental words, such that no target word was presented more than once within a block, and across all of the blocks within an attention condition, each target word was presented exactly once across the 80 trials presented at SNR_{SRM} in each spatial condition and the 80 trials at SNR_{HIGH} in the F-RF condition, and exactly twice across the 160 trials at SNR_{HIGH} in the F-F condition.

2.2.5 EEG Recording and Processing

Electrical Geodesics, Inc. hardware (HydroCel Geodesic Sensor Nets) was used for recording EEG, and ERPLAB software (Lopez-Calderon & Luck, 2014) running in MATLAB (The MathWorks Inc.) was used for creating the ERPs. The procedures and parameters used for data recording (reference, filtering, sampling rate, impedance) and processing (filtering, segmentation, artifact rejection, re-referencing, baseline correction) were identical to those described in Study 1. The resulting ERPs (600-ms duration with 100-ms pre-stimulus baseline) were time-locked to the acoustic onsets of the auditory targets presented in each condition.

2.2.6 Statistical Analysis

2.2.6.1 Behavior

Behavioral performance on the auditory task (Attend Auditory condition) was assessed by calculating the proportion of "yes" responses (number of "yes" responses out of total number of responses) and d' statistic at each experimental SNR in each spatial condition, as described in Study 1. Auditory trials on which participants failed to make a response were excluded from these measures. Behavioral performance on the visual task (Attend Visual condition) was assessed with measures of hit and false-alarm rates as

calculated in prior research under similar conditions (Zobel et al., 2015). The hit rate was defined as the probability of any response being made in the 1350-ms response window 150-1500 ms after the onset of a visual target (either a correct or incorrect button press relative to the target letter case was counted as a hit; the percentage of hits resulting from a correct button press is also reported). The false alarm rate was defined as the probability of any response being made within any other 1350-ms time window within the visual stream.

2.2.6.2 ERPs

Analysis involved the same subset of 81 electrodes arranged in the 3 [Anterior (A), Central (C) Posterior (P)] x 3 [Left (L), Medial (M), Right (R)] grid across the scalp, as described in Study 1. Mean ERP amplitudes were measured across the same four time windows used in Study 1 (P1: 20-60 ms; N1: 130-180 ms; P2: 230-330 ms; P3: 330-500 ms). Visual inspection of the grand-average waveforms also motivated measurement of mean amplitudes in an additional time window for exploratory analysis (N1-late: 200-230 ms). The 50% fractional N1 (70-200 ms) and P2 (200-330 ms) local peak latencies (Kiesel et al., 2008; Lopez-Calderon & Luck, 2014; Luck, 2014) at SNR_{HIGH} were also measured with the jackknife procedure used in Study 1 (Smulders, 2010). To further reduce noise, latency measurements were constrained to the anterior and central electrodes, where the cortical auditory evoked potentials were largest and most apparent in both spatial conditions.

All measurements were collapsed across the 9 electrodes in each cell of the 3x3 scalp grid and entered into repeated-measures ANOVAs in the following order: First, to assess the extent to which the ERP indices of spatial release from informational masking

were replicated from Study 1, analyses of the spatial effects in the Attend Auditory condition were conducted using a whole-scalp repeated measures ANOVA with Spatial Condition (F-F, F-RF), Anterior-Posterior Electrode Position (A, C, P), and Left-Right Electrode Position (L, M, R) entered as separate factors. In cases where a main effect was not observed across the whole scalp, a marginal ($.05 < p \leq .06$) main effect or any interaction between spatial condition and electrode position motivated follow-up analysis on a subset of electrode positions to assess whether a more localized effect was present. Next, identical analyses were conducted in the Attend Visual condition to assess the extent to which spatial effects were observed when listeners attended away from the sounds. Analyses in the Attend Visual condition were constrained to the electrode positions where any spatial effects had been observed in the Attend Auditory condition. Finally, the role of attention in the spatial release from informational masking was assessed by examining the interaction between Spatial Condition and Attention Condition (Attend Auditory, Attend Visual) when both were entered as separate factors in a single repeated measures ANOVA. Again, these analyses were constrained to the electrode positions where any spatial effects had been observed in the Attend Auditory condition. For all analyses, the uncorrected degrees of freedom are reported while the Greenhouse-Geisser correction was applied to the *p*-values.

2.3 Results

2.3.1 Attend Auditory Condition

2.3.1.2 Behavior

Masking thresholds obtained from the preliminary adaptive screening procedure

showed that spatial release from masking (range: 6.63 – 24.63 dB SNR of release) was generally large in the F-RF condition ($M_{threshold}$ = -18.74, SD = 4.33 dB SNR) compared to the F-F condition ($M_{threshold}$ = -1.40, SD = 4.44 dB SNR). Throughout the experimental section, response rates on the auditory task remained high, with no participant failing to respond on more than two trials presented in any Attend Auditory condition or more than four out of the 640 Attend Auditory trials presented across the experiment. Figure 6 shows the mean performance (proportion of "yes" responses and d' values embedded in the bars) on the auditory task in the experimental section. Performance was consistent with the criteria used for choosing the experimental SNRs (Table 1). At SNR_{SRM}, a large release from masking was observed for all participants in the F-RF condition compared to the F-F condition. While performance at SNR_{HIGH} was good in both spatial conditions, a benefit of spatial separation was still observed for all participants in the F-RF condition. All participants also exhibited a decrease in "yes" responses (i.e., false alarms) at SNR_{NULL} in the F-RF condition compared to the F-F condition. A paired-samples t-test found no difference between the proportion of "yes" responses in the F-F condition (M = .50 SD = .07) compared to the F-RF condition (M = .48 SD = .01) across all Attend Auditory trials (p = .24).

2.3.1.3 ERPs

2.3.1.3.1 SNR_{SRM}

Figure 7 shows the grand-average ERPs time-locked to targets presented at SNR_{SRM} in the F-F and F-RF conditions while participants were attending to the auditory task. In the F-F condition, little, if any, effect of target onset was observed in the grand

average. In contrast, identical targets presented in the F-RF condition elicited broadly distributed N1-P2 waveforms (N1 ≈ 170 ms, P2 ≈ 300 ms), consistent with cortical auditory evoked potentials. Additionally, the waveforms in the F-RF condition exhibited negative amplitude across the N1-late time window at anterior-central medial-right electrodes, and a sustained posterior positivity across the P3 time window at central and posterior electrodes. No differences in mean amplitude were found between the F-F and F-RF conditions in the P1 time window (Spatial Condition main effect and interactions with electrode position factors: ps ≥ .49). Targets elicited an N1 in the F-RF condition compared to the F-F condition [$F(1, 17) = 5.22$, $p = .04$, $\eta_p^2 = .24$] that was largest over anterior and central medial electrodes [Spatial Condition x Anterior-Posterior x Left-Right: $F(4, 68) = 10.49$, $p < .001$, $\eta_p^2 = .38$]. In the N1-late time window, a main effect of Spatial Condition was not found across the whole scalp ($p = .08$), but ERPs elicited by targets were more negative in the F-RF condition compared to the F-F condition at anterior-central medial-right electrodes [Spatial Condition x Left-Right interaction: $F(2, 34) = 4.46$, $p = .02$, $\eta_p^2 = .21$; Spatial Condition x Anterior-Posterior x Left-Right interaction: $F(4, 68) = 3.51$, $p = .02$, $\eta_p^2 = .17$; Spatial Condition main effect at anterior-central medial-right electrodes: $F(1, 17) = 5.94$, $p = .03$, $\eta_p^2 = .26$]. Targets elicited a P2 in the F-RF condition compared to the F-F condition [$F(1, 17) = 6.30$, $p = .02$, $\eta_p^2 = .27$], with the largest differences observed at central and medial electrodes [Spatial Condition x Anterior-Posterior x Left-Right: $F(4, 68) = 7.68.$, $p < .001$, $\eta_p^2 = .31$]. In the P3 time window, ERPs were more positive in the F-RF condition compared to the F-F condition [$F(1, 17) = 5.20$, $p = .04$, $\eta_p^2 = .23$], with the largest differences observed at central and posterior electrodes [Spatial Condition x Anterior-Posterior: $F(2, 34) = 7.87$, $p = .009$,

η_p^2 = .32; Spatial Condition x Anterior-Posterior x Left-Right: $F(4, 68) = 8.88$, $p < .001$, η_p^2 = .34].

2.3.1.3.2 SNR_{HIGH}

Figure 8 shows the grand-average ERPs time-locked to targets presented at SNR_{HIGH} in the F-F and F-RF conditions while participants were attending to the auditory task. Cortical auditory evoked potentials elicited by targets presented at SNR_{HIGH} were visible in both spatial conditions (F-F: N1 ≈ 165 ms, P2 ≈ 310 ms ; F-RF: N1 ≈ 155, P2 ≈ 280). In the P1 time window, ERPs elicited by targets were more negative in the F-RF condition compared to the F-F condition [$F(1, 17) = 7.63$, $p = .01$, $\eta_p^2 = .31$]. A main effect of Spatial Condition was not found across the scalp in the N1 time window (p = .21), but N1 amplitude was found to be larger in the F-RF condition compared to the F-F condition at central-posterior left electrodes [Spatial Condition x Anterior-Posterior x Left-Right: $F(4, 68) = 2.67$, $p = .051$, $\eta_p^2 = .14$; Spatial Condition main effect at central-posterior left electrodes: $F(1, 17) = 6.45$, $p = .02$, $\eta_p^2 = .28$]. In the N1-late time window, ERPs were more positive in the F-RF condition compared to the F-F condition across the scalp [$F(1, 17) = 9.97$, $p = .006$, $\eta_p^2 = .37$], with the largest differences observed at central and medial electrodes [$F(4, 68) = 3.05$, $p = .04$, $\eta_p^2 = .15$]. P2 amplitude was larger in the F-RF condition compared to the F-F condition [$F(1, 17) = 13.02$, $p = .002$, $\eta_p^2 = .43$], with the largest differences observed at central and medial electrodes [Spatial Condition x Anterior-Posterior interaction: $F(2, 34) = 5.55$, $p = .02$, $\eta_p^2 = .25$; Spatial Condition x Left-Right interaction: $F(2, 34) = 6.32$, $p = .006$, $\eta_p^2 = .27$; Spatial Condition x Anterior-Posterior x Left-Right interaction: $F(4, 68) = 4.75$, $p = .004$, $\eta_p^2 = .22$]. In the

P3 time window, a main effect of Spatial Condition was not observed across the whole scalp (p = .10), but ERPs were more positive in the F-RF condition compared to the F-F condition at central and posterior electrodes [Spatial Condition x Anterior-Posterior interaction: $F(2, 34) = 4.54$, $p = .03$, $\eta_p^2 = .21$; Spatial Condition x Left-Right interaction: $F(2, 34) = 4.87$, $p = .02$, $\eta_p^2 = .22$; Spatial Condition x Anterior-Posterior x Left-Right interaction: $F(4, 68) = 2.98$, $p = .03$, $\eta_p^2 = .15$; Spatial Condition main effect at central and posterior electrodes: $F(1, 17) = 5.49$, $p = .03$, $\eta_p^2 = .24$]. The 50% fractional peak latencies at anterior and central electrodes were shorter for the N1 [$F(1, 17) = 14.00$, $p = .002$, $\eta_p^2 = .45$] and P2 [$F(1, 17) = 6.59$, $p = .02$, $\eta_p^2 = .28$] in the F-RF condition compared to the F-F condition.

2.3.2 Attend Visual Condition

2.3.2.1 Behavior

Performance on the visual task was characterized by a high hit rate (M = .71, SD = .13) relative to a low false alarm rate (M = .06, SD = .04). Most of the hits (M = 92.65%, SD = 3.16%) resulted from correct button presses in response to the letter cases of the visual targets.

2.3.2.2 ERPs

2.3.2.2.1 SNR_{SRM}

Figure 9 shows the grand-average ERPs time-locked to targets presented at SNR_{SRM} in the F-F and F-RF conditions while participants were attending to the visual task. In the F-F condition, little, if any, effect of target onset was observed in the grand

average. Identical targets presented in the F-RF condition elicited broadly distributed N1-P2 waveforms (N1 ≈ 150 ms, P2 ≈ 230 ms), consistent with cortical auditory evoked potentials. Targets elicited an N1 that was marginal across the scalp in the F-RF condition compared to the F-F condition [$F(1, 17) = 4.04$, $p = .06$, $\eta_p^2 = .19$] and significant over anterior and central medial electrodes [$F(1, 17) = 4.90$, $p = .04$, $\eta_p^2 = .22$]. A separate repeated measures ANOVA with Spatial and Attention conditions entered as separate factors did not show an interaction between Spatial and Attention conditions in the N1 time window either across the whole scalp (two-way and higher-order interactions with electrode positions: ps ≥ .10) or at anterior and central medial electrodes ($p = .99$). In the N1-late time window, a main effect of Spatial Condition was not found at anterior-central medial-right electrodes ($p = .37$) and a marginal interaction between the Spatial and Attention conditions suggested larger differences between the Spatial conditions in the Attend Auditory condition compared to the Attend Visual condition [$F(1, 17) = 4.32$, $p = .053$, $\eta_p^2 = .20$]. Whole-scalp analysis did not show an effect of Spatial Condition on P2 amplitude (Spatial Condition main effect and interactions with electrode positions: ps: ≥ .18). An interaction between Spatial and Attention conditions on P2 amplitude was not evident across whole scalp ($p = .07$) but an interaction was observed at central and medial electrodes showing that Spatial differences were larger in the Attend Auditory condition compared to the Attend Visual condition [Spatial Condition x Attention Condition x Anterior-Posterior x Left-Right: $F(4, 68) = 3.37$, $p = .03$, $\eta_p^2 = .17$; Spatial Condition x Attention Condition at central and medial electrodes: $F(1, 17) = 7.05$, $p = .02$, $\eta_p^2 = .29$]. No effect of Spatial Condition on P3 amplitude was evident across the scalp (Spatial Condition main effect and interactions

with electrode positions: *p*s: ≥ .14) and an interaction between Spatial and Attention conditions showed that P3 differences were larger in the Attend Auditory condition compared to the Attend Visual condition [Spatial Condition x Attention Condition: $F(1, 17) = 7.70$, $p = .01$, $\eta_p^2 = .31$], especially at central and posterior electrodes [Spatial Condition x Attention Condition x Anterior-Posterior: $F(2, 34) = 4.69$, $p = .04$, $\eta_p^2 = .22$; Spatial Condition x Attention Condition x Anterior-Posterior x Left-Right: $F(4, 68) = 3.52$, $p = .02$, $\eta_p^2 = .17$].

2.3.2.2.2 SNR_{HIGH}

Figure 10 shows the grand-average ERPs time-locked to targets presented at SNR_{HIGH} in the F-F and F-RF conditions while participants were attending to the visual task. Cortical auditory evoked potentials elicited by targets presented at SNR_{HIGH} were visible in both spatial conditions (F-F: N1 ≈ 155, P2 ≈ 285; F-RF: N1 ≈ 135 ms, P2 ≈ 260 ms). No effect of Spatial condition on P1 amplitude was observed across the scalp (Spatial Condition main effect and interactions with electrode positions: *p*s: ≥ .51) and an interaction between Spatial and Attention conditions showed that Spatial differences were larger in the Attend Auditory condition compared to the Attend Visual condition [$F(1, 17) = 5.31$, $p = .03$, $\eta_p^2 = .24$]. N1 amplitude was not shown to differ between Spatial conditions at central and posterior left electrodes ($p = .09$), and no interaction between Spatial and Attention conditions was found at those electrodes ($p = .35$). Effects of Spatial Condition on mean amplitudes were not observed across the scalp in the N1-late time window (Spatial Condition main effect and interactions with electrode positions: *p*s ≥ .24), P2 time window (Spatial Condition main effect and interactions with electrode

positions: *p*s ≥ .08) or P3 time window (Spatial Condition main effect and interactions with electrode positions: *p*s ≥ .11). Interactions between Spatial and Attention conditions were evident across the scalp in the N1-late time window [Spatial Condition x Attention Condition: $F(1, 17) = 9.85$, $p = .006$, $\eta_p^2 = .37$], P2 time window [Spatial Condition x Attention Condition: $F(1, 17) = 13.58$, $p = .002$, $\eta_p^2 = .44$; Spatial Condition x Attention Condition x Anterior-Posterior x Left-Right: $F(4, 68) = 3.09$, $p = .04$, $\eta_p^2 = .15$], and P3 time window [$F(1, 17) = 15.01$, $p = .001$, $\eta_p^2 = .47$], showing that amplitude differences between Spatial conditions were larger in the Attend Auditory condition compared to the Attend Visual condition. The 50% fractional peak latencies at anterior and central electrodes were shorter for the N1 [$F(1, 17) = 5.31$, $p = .03$, $\eta_p^2 = .24$] and P2 [$F(1, 17) = 4.66$, $p = .05$, $\eta_p^2 = .22$] in the F-RF condition compared to the F-F condition. No interactions between Spatial and Attention conditions were found at anterior and central electrodes on N1 latency (two-way and higher-order interactions with electrode positions: *p*s ≥ .55) or P2 latency (two-way and higher-order interactions with electrode positions: *p*s ≥ .52).

2.4 Discussion

The present study shed light on the preattentive and attentive processing underlying spatial release from informational masking. Combining the methods in Study 1 with a strong attention manipulation allowed for the comparison of spatial effects on ERPs elicited by identical targets when listeners were attending to the sounds and attending away from the auditory modality. Results showed that within complex listening environments, a spatial cue reduces target/masker confusion by facilitating an

automatic improvement in the early perceptual representation of target sounds and allowing listeners to better direct attention to the target and ignore the masker. Although these results are generally consistent with a two-stage model in which bottom-up processing separates target and masker into independent auditory objects that are then selectively attended (Ihlefeld & Shinn-Cunningham, 2008b, 2008a), the present study revealed tentative evidence that attention may play a more crucial role in the bottom-up processing of sounds that contribute to the spatial release from informational masking.

2.4.1 Attend Auditory Condition

2.4.1.1 Behavior

The behavioral data (Figure 6) showed that participants generally exhibited a sizeable spatial release from informational masking consistent with results from Study 1 (Figure 2) and prior research under similar conditions (Freyman et al., 2008; Morse-Fortier et al., 2017; Zobel et al., 2019). The average masking threshold when the auditory target and masker were spatially co-located was close to the purported limit (0 dB SNR) for informational masking (Arbogast et al., 2005; Freyman et al., 2008), suggesting that non-spatial cues were minimized and that the large benefit observed when target and masker were spatially separated was driven predominantly by the spatial cue. Framing the experiment as a game against the computer kept participants motivated and engaged with the auditory task throughout the Attend Auditory blocks and kept failures to respond on trials to a minimum. Furthermore, the doubling of trials at SNR_{HIGH} in the F-F condition and SNR_{NULL} in the F-RF condition resulted in a near-perfect balance of "yes" and "no" responses within and across spatial conditions, minimizing any effects of

perceived target frequency on the behavioral and ERP results. Consistent with Study 1, the screening procedure was effective at choosing the SNRs according to the experimental criteria (Table 1), such that a strong effect of spatial release from informational masking was exhibited at SNR_{SRM}, and that auditory targets were mostly unmasked in both spatial conditions at SNR_{HIGH}, though some benefit from the spatial cue was still observed. Results again showed that spatially separating the auditory target and masker improved participants' ability to tell both when a target was present (increase in correct "yes" responses) and when a target was absent (decrease in "yes" responses at SNR_{NULL}). As discussed in Study 1, these spatial benefits on behavior suggest that the F-RF condition may have allowed listeners to adopt the simple strategy of responding "yes" to any sound heard from the front, compared to the more difficult task in the F-F condition of deciding whether a fluctuation in sound at the front was produced by the presence of a target or a variation in the masker. Improvements in the ability to tell when an auditory target is present and, importantly, when a target is absent suggests that spatial separation allows listeners to better direct and maintain attention to the target and ignore distractions from the masker.

2.4.1.2 ERPs

2.4.1.2.1 SNR_{SRM}

The ERP effects of virtual separation for auditory targets presented at SNR_{SRM} when participants were attending to the auditory task (Figure 7) replicated the results from Study 1 (Figure 4). Auditory targets at SNR_{SRM} were presented well above the energetic masking threshold (M = 14.50 dB SNR above F-RF threshold, SD_{Diff} = 4.80 dB

SNR). Yet, no apparent effects of target presentation were observed when targets were masked in the F-F condition, suggesting that informational masking can severely disrupt the early perceptual processing of relevant speech sounds. In striking contrast, the same auditory targets elicited broadly distributed cortical auditory evoked potentials when they were unmasked in the F-RF condition, showing that spatial separation can dramatically improve the perceptual representation of relevant sounds. These results again shed light on the early perceptual benefits of spatial separation that help listeners reduce confusion and solve the cocktail party problem under challenging listening conditions. As discussed in Study 1, questions remain about whether these effects reflect modulations of the cortical auditory evoked potentials themselves and whether informational masking and its spatial release may be explained by the degree of overlap between the neural populations responsible for representing relevant and irrelevant sounds (Durlach et al., 2003; Zhang et al., 2014). Further research will be required to answer these questions, including research designed to manipulate energetic and informational masking in both spatial conditions while tracing target representation from the auditory brainstem to the cortex.

In addition to observing the cortical auditory evoked potentials elicited by targets presented at SNR_{SRM} in the F-RF condition, exploratory analysis showed that ERPs were more negative in the N1-late time window when targets were spatially separated compared to co-located with maskers. Effects within this time window were not considered a priori, limiting the ability to draw strong conclusions. This difference may reflect an early component of the negative difference wave (Nd_e) that is often observed for attended vs. unattended stimuli (Luck & Kappenman, 2011; Woods, 1990), insofar as

auditory targets in the present study were better attended when they were spatially separated compared to co-located with maskers. Alternatively, the polarity, timing, and distribution of the effect are consistent with an ORN, and may reflect the perception of the spatially separated target and masker as two independent auditory objects in the F-RF condition, compared to the perception of the co-located target and masker as a single unified auditory object in the F-F condition (Alain et al., 2001, 2002; Alain & Izenberg, 2003; Dyson et al., 2005; Hautus & Johnson, 2005; Zobel et al., 2015). These effects are considered in more detail below when discussing the role of attention in the spatial release from informational masking.

Similar to results from Study 1, a late positivity was observed in the present study when auditory targets presented at SNR_{SRM} were spatially separated compared to co-located with maskers. The timing and polarity of this effect is consistent with a P300 typically elicited by task-relevant stimuli that is attended, such as the target stimuli that participants were asked to detect (Polich, 2007). As discussed in Study 1, the presence of a P300 in response to auditory targets in the F-RF condition suggests that spatial separation allowed listeners to better attend to, recognize, and classify the target sounds in accordance with the auditory task. It is unclear why the P300 effect observed in the present study was more broadly distributed across the hemispheres, while the effect observed in Study 1 was right-lateralized. The tentative hypothesis offered for the right-lateralized effect observed in Study 1 was that the presence of the masker on the right had created a better ear for listening on the left, producing a hemispheric difference similar to those that have been observed in relation to the location of an eliciting stimulus in prior research (Gilmore et al., 2009; Wolpaw & Penry, 1977). Since the auditory target was

always presented from the front in the present studies, however, the effect of a better ear for listening may not have been robust enough to replicate a lateralized P300 across the studies. Further research will be needed to understand the potential relationship between the location of a stimulus and the distribution of ERP responses under the present conditions.

2.4.1.2.2 SNR_{HIGH}

The ERP effects of virtual separation for auditory targets presented at SNR_{HIGH} when participants were attending to the auditory task (Figure 8) were generally consistent with results from Study 1 (Figure 5), though some notable differences were observed. As found in Study 1 and in the prior virtual separation studies (Zhang et al., 2014, 2019), targets presented above the informational masking threshold elicited broadly distributed cortical auditory evoked potentials in both spatial conditions, showing that the SNR alone at this level provided a strong enough cue to reduce target/masker confusion. Results were also generally consistent with an increase in amplitude and decrease in latency of the cortical auditory evoked potentials elicited by auditory targets in the F-RF condition compared to the F-F condition, suggesting that the spatial cue provided some additional benefits to the perceptual representation of targets, consistent with the behavioral results.

In contrast to Study 1, auditory targets presented at SNR_{HIGH} in the present study elicited ERPs that were more negative in the P1 time window when spatially separated compared to co-located with maskers. This result is difficult to explain, since P1 differences were not found in Study 1 and were not assessed in the prior virtual separation studies (Zhang et al., 2014, 2019). It is possible that this P1 negativity reflects a response in the F-RF condition to portions of the target sounds that occurred prior to the

target onset, defined at t = 0 as the first appreciable burst in target energy. Note that the cortical auditory evoked potentials in the F-RF condition exhibited shorter onset latencies, suggesting that earlier portions of the target sounds were unmasked. Since the target onset, as defined, typically occurred near the transition between the first consonant and vowel, targets with longer first consonants contained consonant sound that preceded the time-locked onset. These pre-onset sounds may have occurred at an SNR that was masked in the F-F condition and unmasked in the F-RF condition, eliciting an evoked potential observed in the P1 time window when target and masker were spatially separated. However, it is difficult to reconcile this explanation with the fact that a P1 effect was not found in Study 1, despite having observed shorter N1 and P2 latencies in the F-RF condition in that study. The average SNR_{HIGH} that was chosen for participants (Table 1) and behavioral performance on the auditory task at SNR_{HIGH} were similar in Study 1 and Study 2 (Figures 2 and 6, respectively). However, numerically in Study 2, the average SNR_{HIGH} was higher, performance was better, spatial release was stronger, and twice the number of trials were presented in the F-F condition, which may have contributed to detecting the P1 effect. Another possibility that must be considered is that the P1 effect presently observed was the result of noise. In contrast with Study 1, participants in Study 2 were not in control of advancing the auditory trials once a block began, leaving little opportunity to resolve issues that may have increased electrophysiological noise (e.g., muscle tension). Another source of electrophysiological noise may have come from the visual stream presented while participants engaged in the auditory task. The visual evoked potentials elicited by the onsets of the letters were not time-locked to the auditory events, but likely added a layer of noise throughout the EEG

recording. Steps should be taken to reduce noise in the future use of this experimental design in order to better assess the validity of the P1 effect presently observed. The screening and experimental sections can be split into two separate sessions in order to increase the number of experimental trials that can be collected, and the visual stream can be made to dissolve from one letter to the next to avoid abrupt visual onsets that would elicit visual evoked potentials.

The effects of spatial separation observed at SNR_{HIGH} in the N1, N1-late, and P2 time windows can be best characterized collectively as an increase in amplitude and decrease in onset latency of the cortical auditory evoked potentials elicited by auditory targets when they were spatially separated compared to co-located with maskers. The effect of spatial separation on N1 amplitude was only found to be significant at a small subset of electrodes in the present study, in contrast to the broadly distributed effect found in Study 1. However, a numerical N1 difference can be observed across a broader range of electrodes in the grand average (Figure 8), and considering that N1 amplitude was larger when target and masker were spatially separated in Study 1 and the two prior virtual separation studies (Zhang et al., 2014, 2019), it is reasonable to conclude that the N1 difference presently observed reflects a weak replication of these prior results. After all, variability in the strength of spatial effects can be expected at SNRs above the informational masking threshold, where there is little informational masking to be released across spatial conditions. That said, since the N1 difference presently observed was in the same negative direction as the P1 difference, a downstream influence of the P1 difference cannot be entirely ruled out. In both the N1-late and P2 time windows, amplitudes were more positive in the F-RF condition compared to the F-F condition.

These results are best considered together as redundant measures of a larger P2 (the positivity in the N1-late time window may also reflect the earlier onset latency of the P2) in response to targets when spatially separated compared to co-located with maskers, consistent with Study 1 and the prior studies of virtual separation (Zhang et al., 2014, 2019). Note that a negative difference was not observed at SNR_{HIGH} in the N1-late time window. Insofar as the negative effect of spatial separation observed at SNR_{SRM} indicates processing associated with attention or auditory object perception, it is fitting that no such effect was observed at SNR_{HIGH}, where targets are likely to be relatively easy to attend, and perceived as independent objects in both spatial conditions. In addition to larger amplitudes, the onset latencies of the N1 and P2 were earlier in the F-RF condition compared to the F-F condition. The latency measurement in the N1 time window may be influenced to some extent by the preceding differences observed in the P1 time window. However, given that the onset latency of the P2 was also shorter, and that shorter N1s and P2s were observed in the Attend Visual condition (to be discussed below), and that similar N1 and P2 latency differences were observed in Study 1 and the prior virtual separation studies (Zhang et al., 2014, 2019), it is reasonable to conclude from the present results that the cortical auditory evoked potentials were elicited earlier in time by auditory targets presented at SNR_{HIGH} when they were spatially separated compared to co-located with maskers. As discussed in Study 1, since the ramping up in energy at the target onset will cross the F-RF threshold before it crosses the F-F threshold, this latency difference may be explained by the spatial unmasking of earlier portions of the target sound in the F-RF condition.

In contrast to Study 1, a late positivity was elicited by auditory targets presented at SNR_{HIGH} when they were spatially separated compared to co-located with maskers. This effect was similar in timing and distribution to the P300 observed at SNR_{SRM} (Polich, 2007). The presence of a P300 at SNR_{HIGH} suggests that although the intensity cue allowed participants to successfully perform the auditory task in the F-F condition, the spatial cue in the F-RF condition provided additional benefits that improved listeners' ability to detect and classify target sounds. It is unclear why a P300 was observed in Study 2 and not Study 1 at SNR_{HIGH}, even though the behavioral and ERP data in both studies showed that some informational masking was spatially released. Again, it may be the case that spatial effects are generally more variable when, on average, there is little informational masking to be released across spatial conditions.

2.4.2 Attend Visual Condition

2.4.2.1 Behavior

Despite the challenging nature of the two-back visual task, results showed that participants were motivated by the game against the computer, remaining actively engaged in the task and successful at detecting the visual targets. Most of the responses made by participants followed the presentation of a visual target and used the correct button press relative to the target's letter case, while importantly, performance did not approach ceiling. This suggests that the visual task remained challenging and required participants to maintain their attention on the task throughout the Attend Visual blocks.

2.4.2.2 ERPs

2.4.2.2.1 SNR_{SRM}

The ERPs elicited by auditory targets presented at SNR_{SRM} in the Attend Visual condition provide a striking account of the bottom up, preattentive processing that underlies spatial release from informational masking. When participants were attending to the visual task, a dramatic modulation of the cortical auditory evoked potentials elicited by auditory targets was observed in the F-RF condition compared to the F-F condition, similar to the pattern of early effects observed in the Attend Auditory condition. Specifically, cortical auditory evoked potentials were elicited by auditory targets when they were spatially separated from maskers, while little to no effect of target presentation was observed when the same targets were co-located with maskers. These results provide compelling evidence that within complex, noisy environments, a spatial cue can reduce target/masker confusion by facilitating preattentive processes that dramatically improve the neural representation of target sounds in the early stages of auditory perception. These findings are consistent with prior studies of virtual separation (Zhang et al., 2014, 2019), but the extent to which preattentive processing contributed to the spatial release from informational masking in the present study has not been previously reported. As such, these ERP results provide strong support for the role of primitive processes in the spatial release from informational masking that Ihlefeld and Shinn-Cunningham (2008b, 2008a) had posited in their model but were only able to tentatively support with their behavioral results. Specifically, Ihlefeld and Shinn-Cunningham (2008b, 2008a) suggested that spatial separation facilitates automatic processes involved in auditory object formation, including the grouping of target sounds

into a separate stream. In the present study, the modulation of the cortical auditory evoked potentials observed across spatial conditions when listeners were attending to the visual task shows that the target sounds were automatically separated from the masker and independently processed when the spatial cue was introduced.

The effects of spatial separation at SNR_{SRM} in the N1-late, P2, and P3 time windows that were observed in the Attend Auditory condition were not observed in the Attend Visual condition. These differences suggest that attention is important for beneficial processing within these time windows and will be addressed below when discussing the role of attention in the spatial release from informational masking.

2.4.2.2.2 SNR_{HIGH}

Cortical auditory evoked potentials were elicited by auditory targets presented at SNR_{HIGH} in both spatial conditions in the Attend Visual condition, consistent with results observed in the Attend Auditory condition, and all attention conditions in the prior virtual separation studies (Zhang et al., 2014, 2019). This result shows that when target sounds are presented above the informational masking threshold, the SNR alone can provide a cue that facilitates their automatic grouping and separation from the masker. Effects of spatial separation on ERP amplitude were not observed at SNR_{HIGH} in any time window when listeners were directing their attention to the visual task. This result differs somewhat from prior virtual separation research reporting larger N1s elicited by auditory targets that were spatially separated from maskers both when listeners directed attention to the target sounds and directed attention away from the target sounds (Zhang et al., 2019). In the present study, however, only a weak effect of spatial separation on N1 amplitude was observed in the Attend Auditory condition at SNR_{HIGH} (for reasons that

are not clear), and no interaction between the spatial and attention conditions was found. Thus, the present results are consistent with the prior research, insofar as attention was not shown to modulate amplitude effects within the N1 time window. The failure to observe spatial effects on P2 amplitude when listeners directed attention away from the target sounds in the present study is also consistent with prior research on virtual separation (Zhang et al., 2019). These results, along with differences found in the P1 and P3 time windows, will be addressed in the section below on the role of attention in the spatial release from informational masking.

When attention was directed to the visual task, the cortical auditory evoked potentials elicited by auditory targets presented at SNR_{HIGH} began earlier in time when targets were spatially separated compared to co-located with maskers. This result is consistent with latency effects observed when listeners were attending to the auditory task in Study 1 and 2, and similar to latency effects observed in prior virtual separation studies (Zhang et al., 2014, 2019). As discussed previously, the shorter latencies observed in the F-RF condition may be attributed to the spatial unmasking of earlier portions of the target onsets. Thus, shorter latencies observed in the Attend Visual condition show that spatial release from informational masking still occurred when participants were attending away from the target sounds. These results provide corroborating evidence that under challenging listening conditions, spatial cues can reduce informational masking by facilitating automatic, bottom-up processes that improve the perceptual representation of relevant sounds.

2.4.3 The Role of Attention in the Spatial Release from Informational Masking

The role of attention in the spatial release from informational masking can be assessed by comparing the ERP effects of spatial separation in the Attend Auditory and Attend Visual conditions (Figures 11 and 12). In the present study, the effects of spatial separation on amplitude in the N1 time window (SNR_{SRM} and SNR_{HIGH}) and the onset latency of the N1 (SNR_{HIGH}) were not shown to differ by whether participants attended to or away from the target sounds (i.e., no interaction between spatial and attention conditions was found). Prior research on virtual separation also failed to show a difference in N1 amplitude and latency effects between attention conditions (Zhang et al., 2019). Together, these results suggest that attention may play a minimal role in the earliest perceptual stages of spatial release from informational masking, and that within complex, noisy environments, early benefits from spatial separation are largely driven by preattentive, bottom-up processing that automatically improves the neural representation of relevant speech sounds. In contrast with results observed in the N1 time window, attention was shown to modulate the effects of spatial separation on P2 and P3 amplitude at SNR_{SMR} and SNR_{HIGH}. These results are consistent with prior virtual separation research reporting similar attentional modulation of spatial effects on P2 amplitude (Zhang et al., 2019). It is not surprising to find that attention plays a crucial role in these later time windows. Research suggests that the P2 is not merely an obligatory component of the cortical auditory evoked potential, but may also reflect higher-level cognitive processes sensitive to training (Tremblay et al., 2014) and crucial for the task-relevant processing associated with the P300 (Crowley & Colrain, 2004), a component that has been strongly linked with the engagement of working memory and attention (Polich,

2007). Simply put, effects within these later time windows are likely to reflect task-relevant cognitive processing involving learning, sound classification, and response selection, that require attentional awareness. Taken together, results in the early (N1) and later (P2, P3) time windows provide some support for Ihlefeld and Shinn-Cunningham's (2008b, 2008a) two-stage model in which the spatial cue facilitates primitive mechanisms that automatically group target and masker sounds into separate auditory objects that listeners can then selectively attend to further improve target processing. However, there are some surprising results from the present study that prevent a clear delineation to be drawn between bottom-up and top-down processing in the spatial release from informational masking, and provide some tentative evidence that attention may play a more crucial role in the early perceptual benefits of spatial separation. Namely, attentional modulation of spatial effects was found in the P1 time window at SNR_{HIGH}, and indicated in the N1-late time window at SNR_{SRM}.

ERP amplitude in the P1 time window was more negative when auditory targets presented at SNR_{HIGH} were spatially separated compared to co-located with maskers in the Attend Auditory condition. Furthermore, attention was shown to modulate this P1 effect such that it was observed when listeners were attending to the sounds, but not observed when attention was directed to the visual task. As discussed above, explanations of this surprising result must be offered with caution, considering that P1 effects were not found in Study 1 and were not analysed in prior ERP studies of virtual separation (Zhang et al., 2014, 2019), and that potential effects of noise cannot be ruled out. It is possible, however, that the negative amplitude in the P1 time window reflects the processing of consonant sounds that preceded the bursts in target energy that

were identified for time-locking the ERPs. That is, at SNR_{HIGH}, the consonant sounds that preceded the defined target onsets were presented at an SNR that allowed them to be spatially released from informational masking. Thus, the P1 effect may have been driven by low-intensity sounds that were presented much closer to the energetic masking threshold (i.e., the F-RF masking threshold). If this is the case, then the attentional modulation of the P1 effect would suggest that attention plays a critical role in the early perceptual stages of spatial release from informational masking when relevant sounds are generally difficult to detect in the presence of a spatial cue. This hypothesis can be tested in the future by presenting listeners with a broader range of SNRs that include lower-intensity targets presented closer to the energetic masking threshold. Furthermore, the present and prior ERP studies of virtual separation used very strong spatial cues (55° and 180° of horizontal separation between target and masker in the present and prior studies, respectively). Small spatial separations between targets and maskers can be used in future investigations to assess the role of attention when listeners are challenged by weaker spatial cues. If attention is shown to modulate early perceptual effects of spatial separation for targets presented closer to the energetic masking threshold and/or closer to the spatial location of the masker, such results would suggest that under more demanding listening conditions, top-down selective attention is required to resolve ambiguities and facilitate the bottom-up representation of relevant speech sounds in the early perceptual stages of spatial release from informational masking.

Effects observed in the N1-late time window for auditory targets presented at SNR_{SRM} may also indicate an important role for attention in the perceptual representation of an auditory scene when competing speech sounds are spatially separated. Auditory

targets that were spatially separated from maskers elicited ERPs that were more negative in the N1-late time window when listeners were attending to the auditory task, while no effect was observed when listeners directed attention to the visual task. Again, only speculative conclusions can be offered for these results, because the analysis was exploratory and the interaction between the spatial and attention conditions was marginal. One possible explanation is that the negativity observed in the N1-late time window in the Attend Auditory condition reflects an Nd_e often observed for attended compared to unattended sounds (Luck & Kappenman, 2011; Woods, 1990). If this is the case, then the marginal interaction between the spatial and attention conditions would indicate that in the Attend Auditory condition, the spatial cue allowed listeners to better attend to the target and ignore the masker, improving the perceptual representation of the target and contributing to a release from informational masking. Furthermore, the fact that no such effects were observed at SNR_{HIGH} would indicate that either informational masking was sufficiently released by bottom-up processing at higher target intensities without demanding additional contributions from attention, or that the amount of masking release was too small for attentional effects to be observed. It is unclear why an attentional effect at SNR_{SRM} would first appear in the N1-late time window and not in the N1 time window in the present study. The latency of early attention effects can differ based on a variety of factors, including the strength and focus of listener attention, task demands, and features of the stimuli and their presentation (Luck & Kappenman, 2011; Woods, 1990). It may be the case in the present study that presenting the visual stream and auditory targets from a similar location in front of the participant may have weakened the attention manipulation, reducing any effect in the N1 time window. Although steps were

taken to avoid the time locking of the auditory and visual onsets, since participants were spatially directing attention to a similar location in both attention conditions, it may have been difficult to selectively attend to only one perceptual modality at a time (i.e., attend only to the visual information while ignoring the auditory information and/or vise versa). Furthermore, although the visual and auditory stimuli were not presented at exactly the same location, it is possible that the visual stimuli captured the auditory stimuli to some extent through a ventriloquist effect, especially when participants were engaged in the visual task (Bruns, 2019). A larger spatial separation between the auditory and visual stimuli can be used in the future to strengthen the attention manipulation and better assess early attentional effects. It is also possible that the auditory task in the present study did not require a strong enough engagement of attention to drive N1 effects. The auditory task was relatively easy at SNR_{SRM} in the F-RF condition and at SNR_{HIGH} in both spatial conditions and may not have sufficiently taxed attentional resources. As discussed above, effects observed in the P1 time window at SNR_{HIGH} in the present study may suggest that attention plays an active role in the early perceptual stages of spatial release from informational masking under more challenging listening conditions in which stimuli are presented closer to the energetic masking threshold. It is also possible that the large P2 and P3 effects observed in the present study overwhelmed a smaller N1 attention effect that would have otherwise been observed. The prior ERP research on virtual separation avoided large late positive effects by using a stimulus to obtain the ERPs that differed from the task-relevant stimulus participants were asked to detect (Zhang et al., 2014, 2019), though attention was still not shown to modulate the effects of spatial separation N1 amplitude (only a main effect of attention was evident). That said, it may

be beneficial to incorporate a similar method into the present paradigm in the future—in addition to the methods described above to reduce electrophysiological noise—to better isolate effects in early time windows. Future research can vary such factors and others that are known to modulate early attention effects to better assess the negativity observed in the N1-late time window and the extent to which attention may influence spatial release from informational masking at different stages of auditory processing.

Another possible explanation for the effect observed in the N1-late time window at SNR_{SRM} in the Attend Auditory condition is that it reflects an ORN elicited by the perception of the target and masker as two auditory objects when spatially separated compared to one unified object when spatially co-located. This would provide corroborating evidence that spatial separation reduces informational masking by improving the perceptual representation of auditory objects within a complex scene. This would also explain why a similar effect was not observed at SNR_{HIGH}, where target and masker would be generally perceived as separate auditory objects in both spatial conditions. If the effect observed at SNR_{SRM} in the Attend Auditory condition is, indeed, an ORN, then the marginal interaction between the spatial and attention conditions has two important implications. First, it would suggest that attention is critical for the beneficial representation of auditory objects in the spatial release from informational masking. Research suggests that the ORN can be elicited automatically under simple conditions of object formation, such as those involving harmonic grouping or dichotic pitch (Alain et al., 2001, 2002; Alain & Izenberg, 2003; Bregman, 1990; Dyson et al., 2005; Hautus & Johnson, 2005). Under more complex conditions, however, involving the grouping of lead and lag sounds in the precedence effect, Zobel et al. (2015) found

that attending to the sounds was required to observe the ORN. Attention may have played a similar critical role within the challenging conditions of the present study, where virtual separation required complex groupings of sounds involving the precedence effect. Given that informational masking is likely to arise under complex listening conditions, it may be the case that listeners often rely upon attention to improve the perceptual representation of auditory objects and reduce confusion. Second, if results at SNR_{SRM} do reflect an attentional modulation of an ORN, it would suggest that the ORN indexes only a perceptual component or outcome associated with auditory object processing. An ORN was not observed when listeners attended to the visual task in the present study, yet auditory evoked potentials were still elicited by auditory targets that were spatially separated compared to co-located with maskers. This would suggest that some processing associated with auditory object formation had separated the target sounds from the masker sounds and grouped them into an individual unit without eliciting an ORN. The ORN would therefore reflect only a particular aspect of auditory object formation, perhaps corresponding to the perceived number of auditory objects or the situating of multiple objects within an auditory scene. More research is needed to fully characterize the processes associated with the ORN, their relationship to attention, and their potential contributions to the spatial release from informational masking.

In summary, results from the present study showed strong contributions from preattentive bottom-up processing in the early perceptual stages of spatial release from informational masking, while the role of attention was most evident at later stages associated with the cognitive processing of relevant sounds. Effects observed in the P1 and N1-late time windows, however, hint that attention may play a more critical role in

the early perceptual benefits of spatial separation when listeners are faced with more challenging conditions in which the SNR may be generally poor or cues may be weak. Further investigation under a variety of conditions designed to challenge the listener is needed to better characterize the relative contributions and potential interdependencies of bottom-up and top-down processing in the spatial release from informational masking.

CHAPTER 3

CONCLUSION

The present studies examined the stages of processing and underlying preattentive and attentive mechanisms involved in the spatial release from informational masking. The ERP results provide compelling evidence that within complex listening environments, spatially separating competing speech streams can elicit strong contributions from automatic, bottom-up processes that reduces confusion and improve the perceptual representation of relevant speech in early stages of auditory processing. Results also show that attention plays a necessary role in supporting higher-level task-relevant cognitive benefits of spatial separation that are important for sound classification at later stages of processing. These results provide strong support for models of spatial release from informational masking that include both bottom-up and top-down contributions. However, Study 2 also revealed new, tentative evidence that attention may benefit processing in earlier perceptual stages of spatial release from informational masking under some conditions. Attentional modulation of spatial effects were observed in earlier time windows (P1 and N1-late time windows) than previously reported (Zhang et al., 2014, 2019), and merits further investigation. It may be the case that within more challenging listening environments, in which relevant sounds are generally difficult to detect and distinguishing cues are weak, attention may play a crucial role in the perceptual representation of relevant sounds and the formation of auditory objects that contribute to the spatial release from informational masking.

Care should be taken not to generalize results from the present studies to all situations in which informational masking might arise. In the present studies, large

benefits were driven automatically by the spatial cue at early stages of perceptual processing, but this result is likely to be related to the fact that the type of informational masking and the type of auditory task (i.e., a target detection task compared to a word identification task) were largely perceptual in nature. Research suggests that informational masking can arise at different stages of processing depending on the context, which may change the stages of processing at which it is spatially released and the relative contributions of bottom-up and top-down processing. For example, informational masking can arise at semantic levels of linguistic processing, such that greater masking occurs when target and masker contain higher- compared to lower-frequency word phrases (Cherry, 1953) or when target and masker both match the native language of the listener compared to when the masker is of a different language (Freyman et al., 2001). In these cases, strong contributions from selective attention may be required to spatially release informational masking at later stages of processing, where effects of attention were most apparent in the present studies (P2 and P3 time windows) and where ERPs often reflect differences associated with semantic processing (e.g., N400 effects) (Kutas & Federmeier, 2011). The present studies examined the effects of spatial separation on a type of informational masking that arises in early perceptual stages of auditory processing. The results establish an effective paradigm and important ERP measures that may be applied to develop a more comprehensive account of the various contexts that produce informational masking and the various mechanisms that contribute to its spatial release. Such research will be important for understanding how listeners solve the cocktail party problem and for designing interventions that benefit those who experience difficulties within complex, challenging listening environments.

TABLES AND FIGURES

Table 1: Description of the Experimental SNRs

Name	Criteria for choosing experimental SNRs for each participant	Mean chosen SNR (*SD*)	
		Study 1	Study 2
SNR_{LOW}	SNR just low enough for Targets to remain undetected in F-RF condition	-25.50 dB (5.21 dB)	
SNR_{SRM}	Highest SNR for spatial release from masking (SRM), such that Targets are undetected in F-F and detected in the F-RF condition	-2.40 dB (1.39 dB)	-4.11 dB (2.08 dB)
$SNR_{SRM-2dB}$	2-dB SNR below SNR_{SRM}		
$SNR_{SRM+2dB}$	2-dB SNR above SNR_{SRM}		
SNR_{HIGH}	SNR just high enough for Targets to be detected in the F-F	+6.35 dB (1.90 dB)	+6.56 dB (1.15 dB)
SNR_{NULL}	No Target presented on a trial		

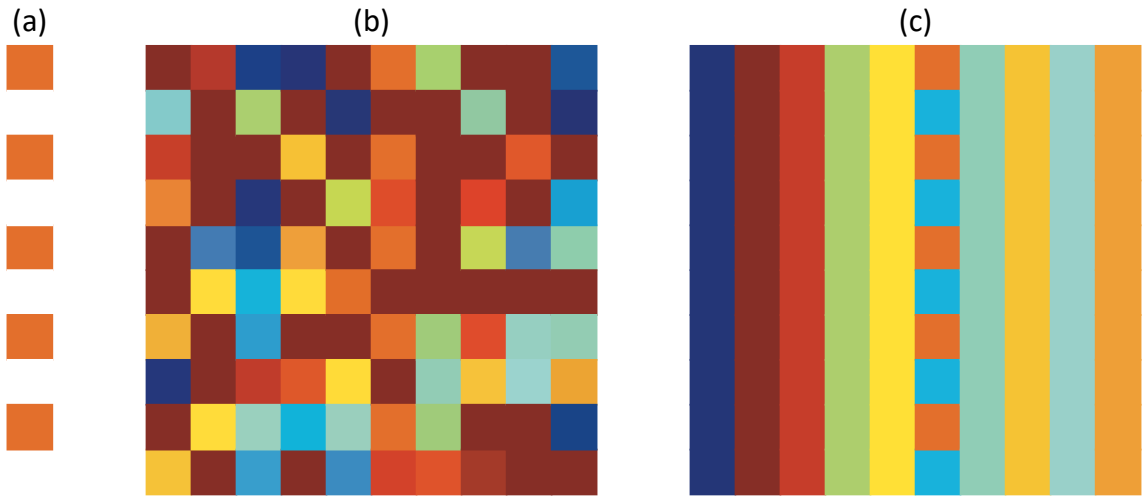

Figure 1: Visual illustration of informational masking. The target (a) is masked in (b) because of perceptual similarity and the lack of a predictable spatial pattern. Informational masking is released in (c). Reproduced from "The Information-divergence Hypothesis of Informational Masking", by R. A. Lutfi, L. Gilbertson, I. Heo, A.-C. Chang, J. Stamas, 2013, *J. Acoust. Soc. Am.*, *134*(3), p. 2161, with the permission of the Acoustical Society of America. Copyright 2013 by the Acoustical Society of America.

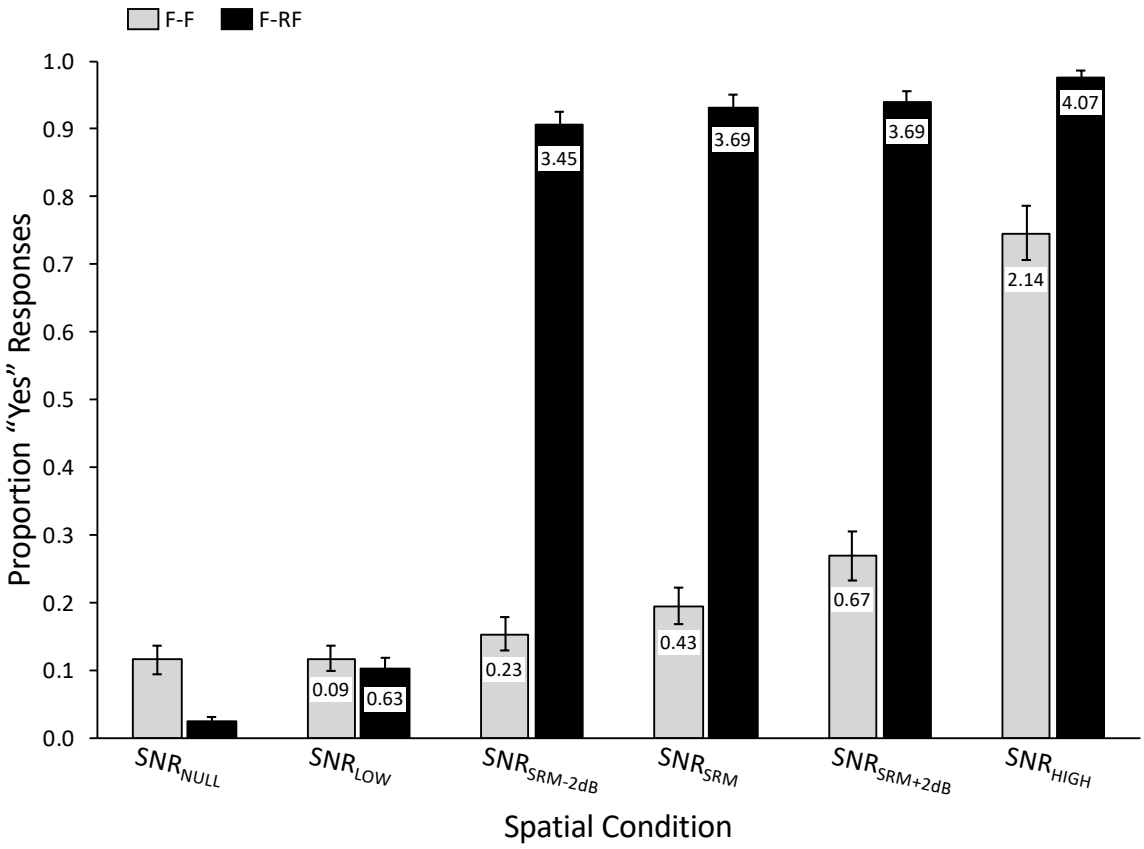

Figure 2: Mean proportion of trials on which the targets were reported to be heard in Study 1 ($N = 20$). Error bars indicate ±1 *SEM;* embedded numbers indicate d'. Spatial release from masking is evident when the response rate or d' is higher in the F-RF condition than in the F-F condition (except at SNR_{NULL}). The false alarm rate for all d' calculations was the response rate at SNR_{NULL} in the respective spatial condition.

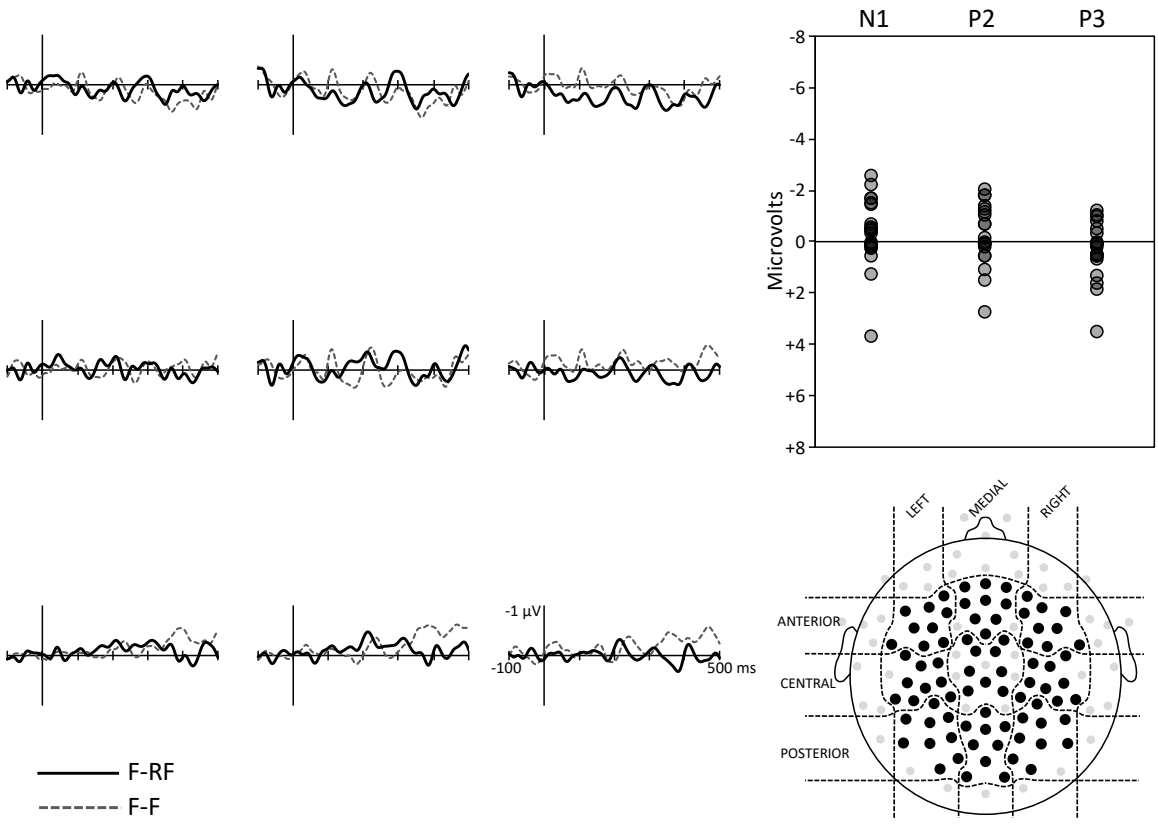

Figure 3: Grand average ERPs elicited by targets at SNR_{LOW} in the F-F and F-RF conditions in Study 1 (*N* = 20). Waveforms are averaged across the nine electrodes in a scalp region indicated in the head map and time locked to the earliest abrupt increase in target amplitude (time = 0 ms). A 30-Hz zero-phase low-pass filter was applied to the data shown, but was not part of analysis. In the upper right, the difference in mean amplitude between the two spatial conditions (F-RF minus F-F) for each individual subject is shown for the N1 (130-180 ms), P2 (230-330 ms), and P3 (330-500 ms) time windows.

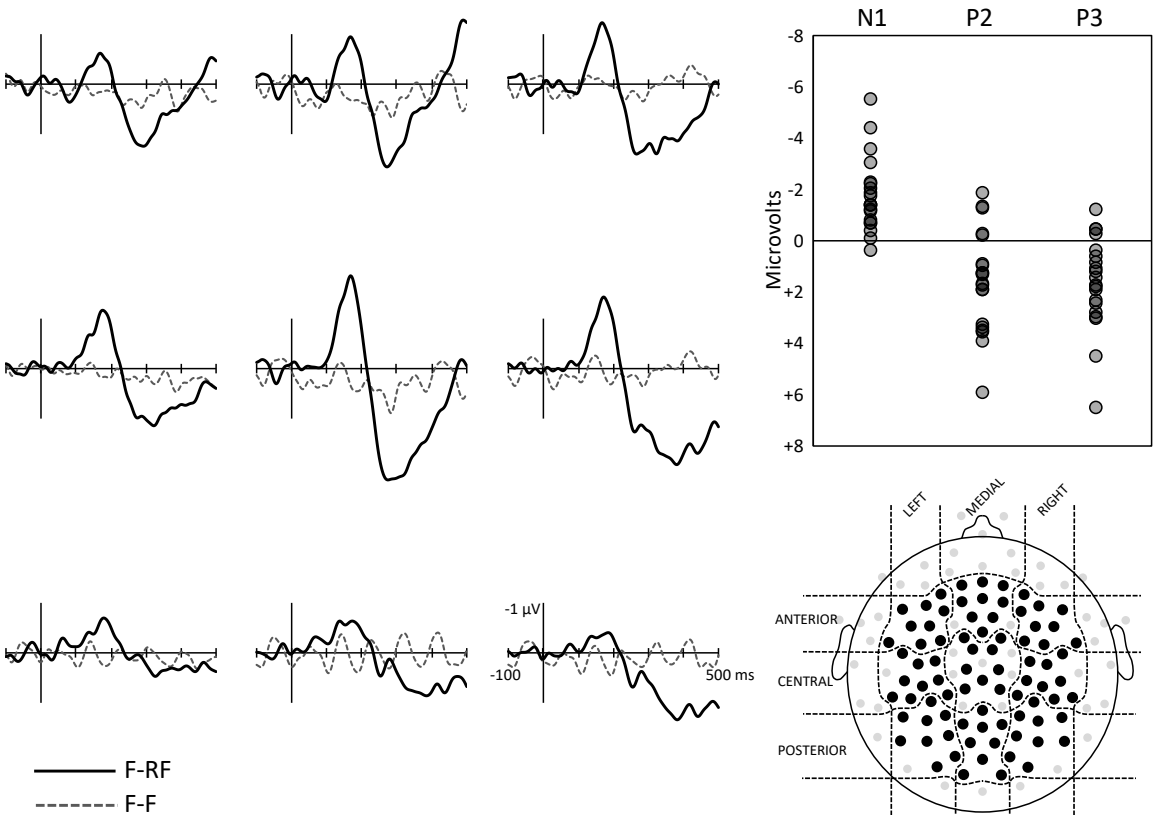

Figure 4: Grand average ERPs elicited by targets at SNR_{SRM} in the F-F and F-RF conditions in Study 1 ($N = 20$). Waveforms are averaged across the nine electrodes in a scalp region indicated in the head map and time locked to the earliest abrupt increase in target amplitude (time = 0 ms). A 30-Hz zero-phase low-pass filter was applied to the data shown, but was not part of analysis. In the upper right, the difference in mean amplitude between the two spatial conditions (F-RF minus F-F) for each individual subject is shown for the N1 (130-180 ms), P2 (230-330 ms), and P3 (330-500 ms) time windows.

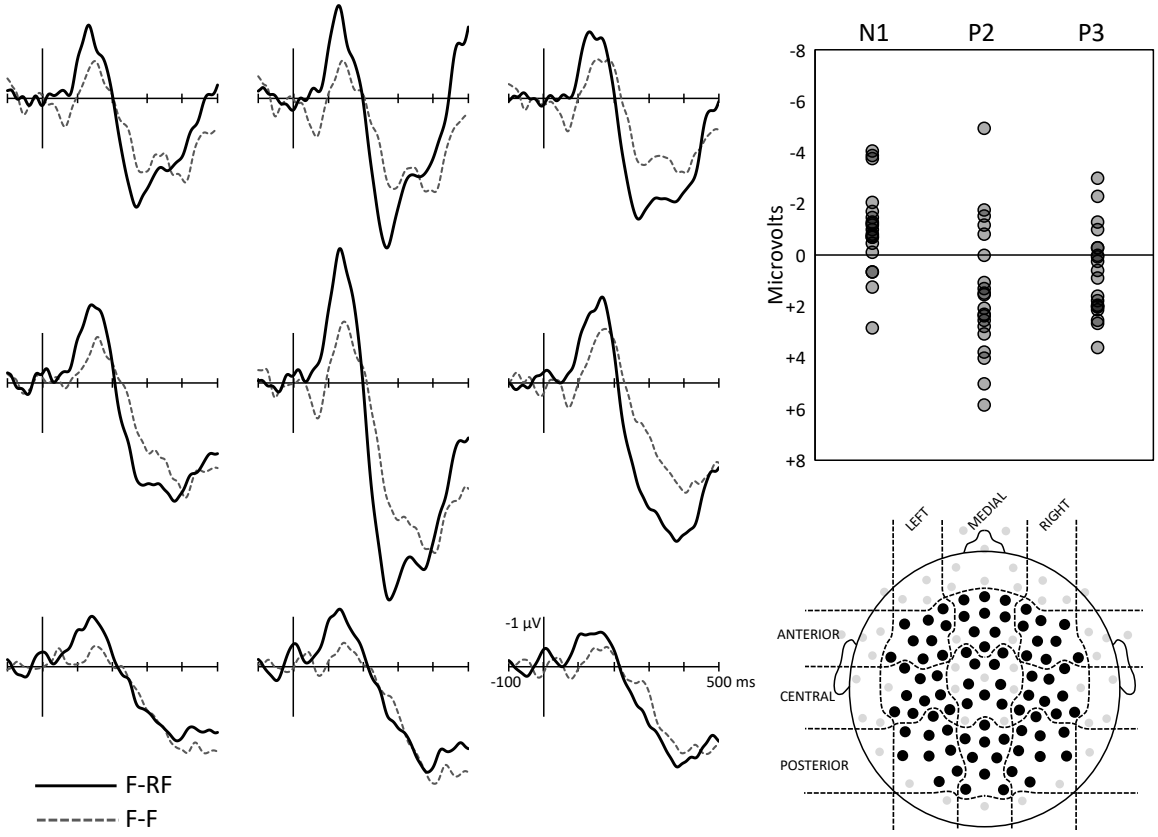

Figure 5: Grand average ERPs elicited by targets at SNR_{HIGH} in the F-F and F-RF conditions in Study 1 ($N = 20$). Waveforms are averaged across the nine electrodes in a scalp region indicated in the head map and time locked to the earliest abrupt increase in target amplitude (time = 0 ms). A 30-Hz zero-phase low-pass filter was applied to the data shown, but was not part of analysis. In the upper right, the difference in mean amplitude between the two spatial conditions (F-RF minus F-F) for each individual subject is shown for the N1 (130-180 ms), P2 (230-330 ms), and P3 (330-500 ms) time windows.

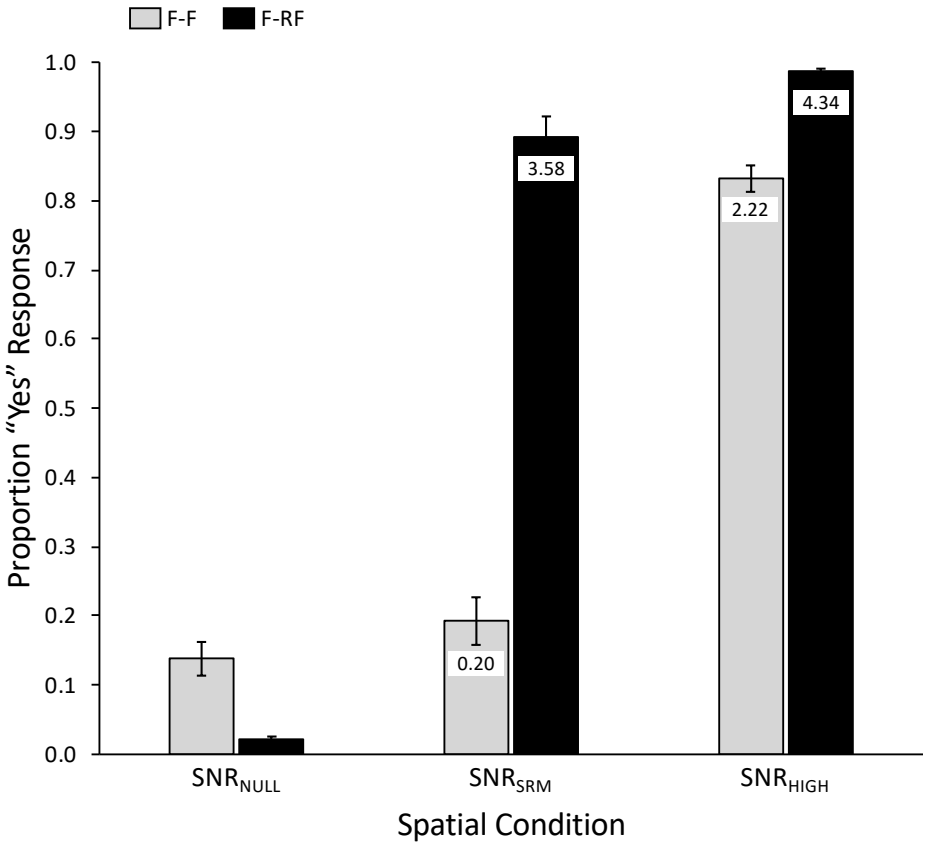

Figure 6: Mean proportion of trials on which the targets were reported to be heard in Study 2 ($N = 18$). Error bars indicate ±1 *SEM;* embedded numbers indicate d'. Spatial release from masking is evident when the response rate or d' is higher in the F-RF condition than in the F-F condition (except at SNR_{NULL}). The false alarm rate for all d' calculations was the response rate at SNR_{NULL} in the respective spatial condition.

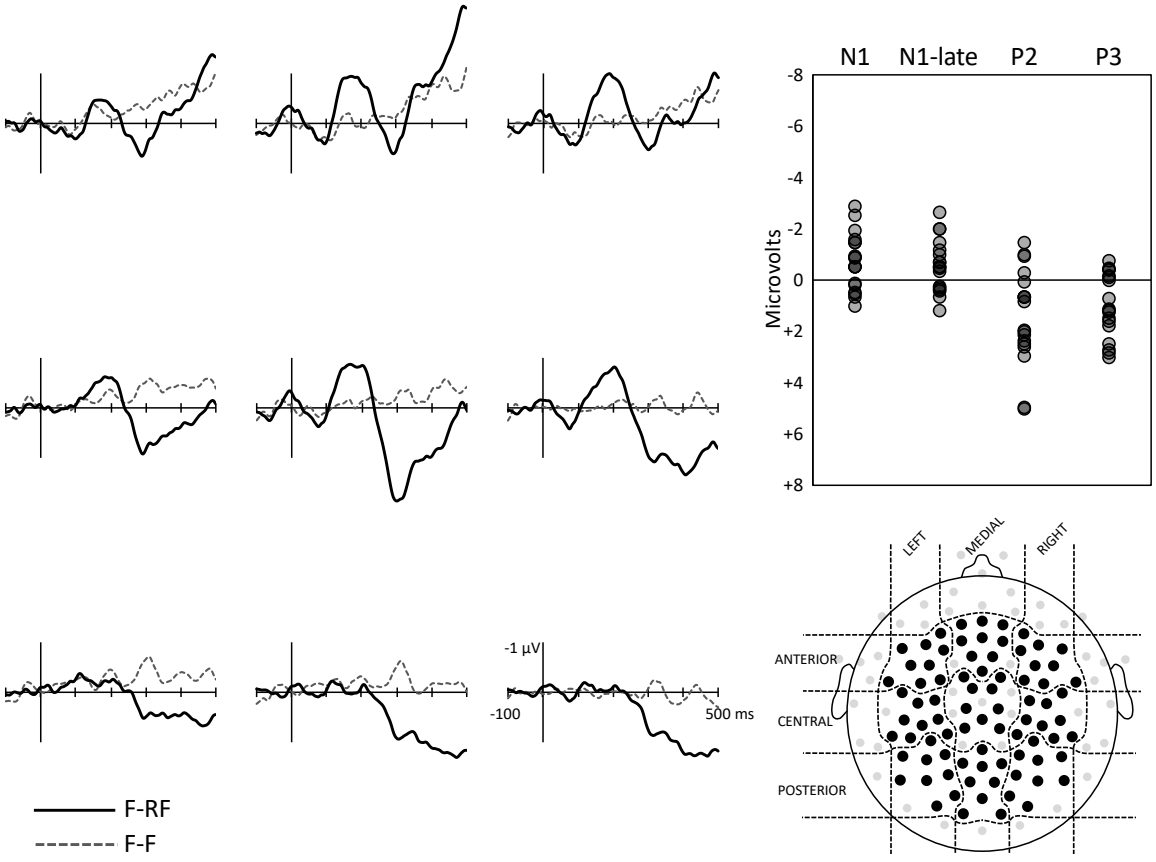

Figure 7: Grand average ERPs elicited by targets at SNR_{SRM} in the F-F and F-RF conditions in the Attend Auditory condition in Study 2 (N = 18). Waveforms are averaged across the nine electrodes in a scalp region indicated in the head map and time locked to the earliest abrupt increase in target amplitude (time = 0 ms). A 30-Hz zero-phase low-pass filter was applied to the data shown, but was not part of analysis. In the upper right, the difference in mean amplitude between the two spatial conditions (F-RF minus F-F) for each individual subject is shown for the N1 (130-180 ms), N1-late (200-230 ms), P2 (230-330 ms), and P3 (330-500 ms) time windows.

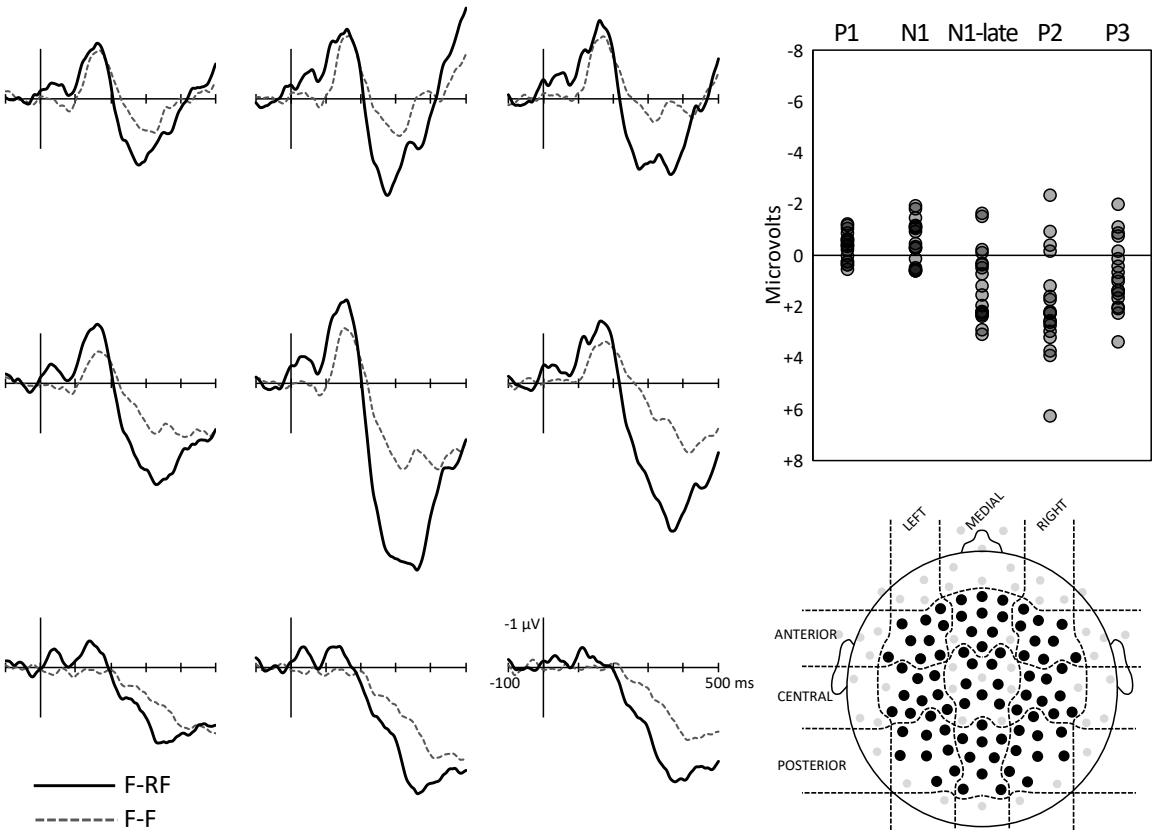

Figure 8: Grand average ERPs elicited by targets at SNR_{HIGH} in the F-F and F-RF conditions in the Attend Auditory condition in Study 2 (N = 18). Waveforms are averaged across the nine electrodes in a scalp region indicated in the head map and time locked to the earliest abrupt increase in target amplitude (time = 0 ms). A 30-Hz zero-phase low-pass filter was applied to the data shown, but was not part of analysis. In the upper right, the difference in mean amplitude between the two spatial conditions (F-RF minus F-F) for each individual subject is shown for the P1 (20-60 ms), N1 (130-180 ms), N1-late (200-230 ms), P2 (230-330 ms), and P3 (330-500 ms) time windows.

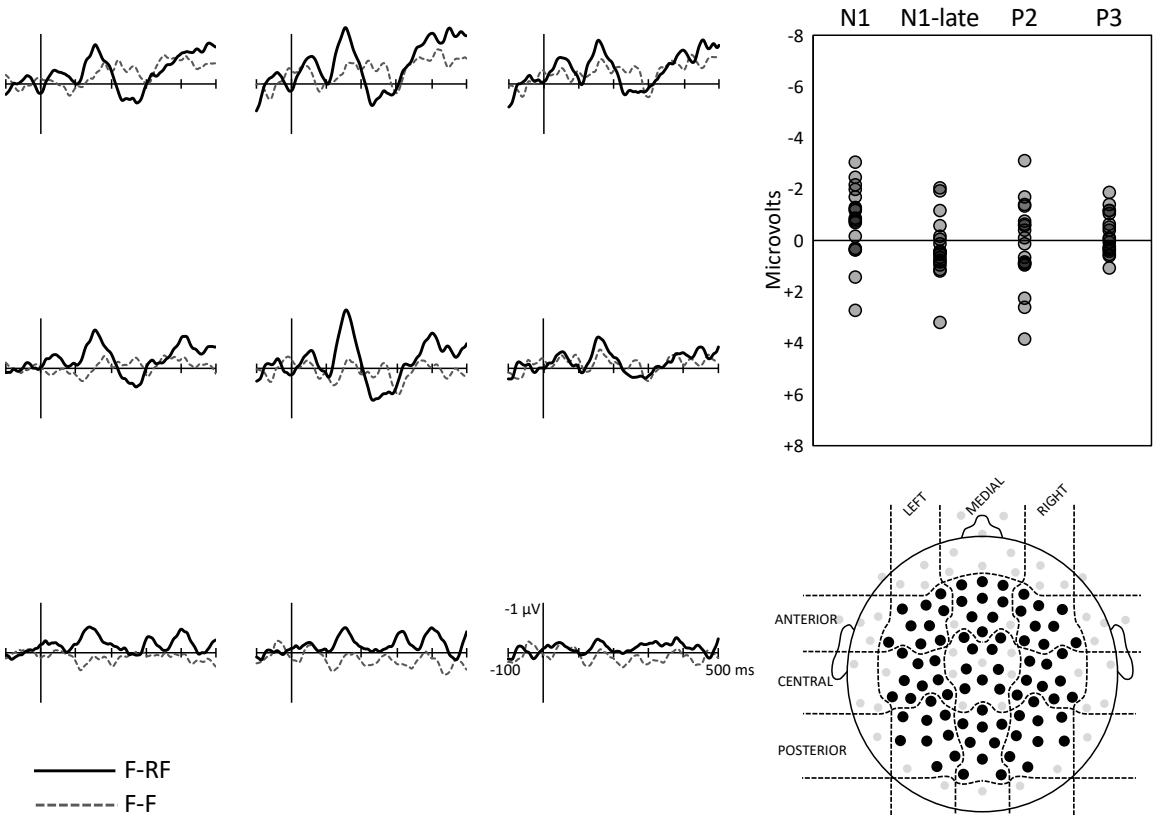

Figure 9: Grand average ERPs elicited by targets at SNR_{SRM} in the F-F and F-RF conditions in the Attend Visual condition in Study 2 (*N* = 18). Waveforms are averaged across the nine electrodes in a scalp region indicated in the head map and time locked to the earliest abrupt increase in target amplitude (time = 0 ms). A 30-Hz zero-phase low-pass filter was applied to the data shown, but was not part of analysis. In the upper right, the difference in mean amplitude between the two spatial conditions (F-RF minus F-F) for each individual subject is shown for the N1 (130-180 ms), N1-late (200-230 ms), P2 (230-330 ms), and P3 (330-500 ms) time windows.

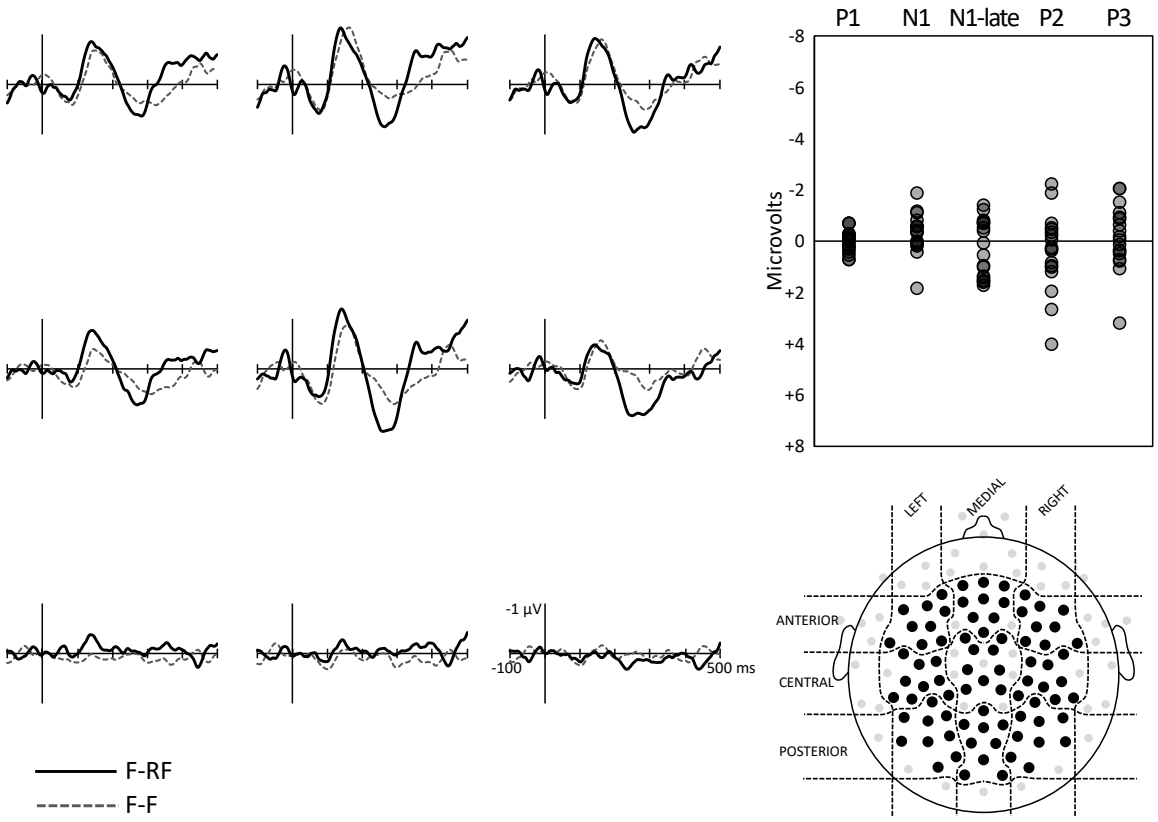

Figure 10: Grand average ERPs elicited by targets at SNR_{HIGH} in the F-F and F-RF conditions in the Attend Visual condition in Study 2 (*N* = 18). Waveforms are averaged across the nine electrodes in a scalp region indicated in the head map and time locked to the earliest abrupt increase in target amplitude (time = 0 ms). A 30-Hz zero-phase low-pass filter was applied to the data shown, but was not part of analysis. In the upper right, the difference in mean amplitude between the two spatial conditions (F-RF minus F-F) for each individual subject is shown for the P1 (20-60 ms), N1 (130-180 ms), N1-late (200-230 ms), P2 (230-330 ms), and P3 (330-500 ms) time windows.

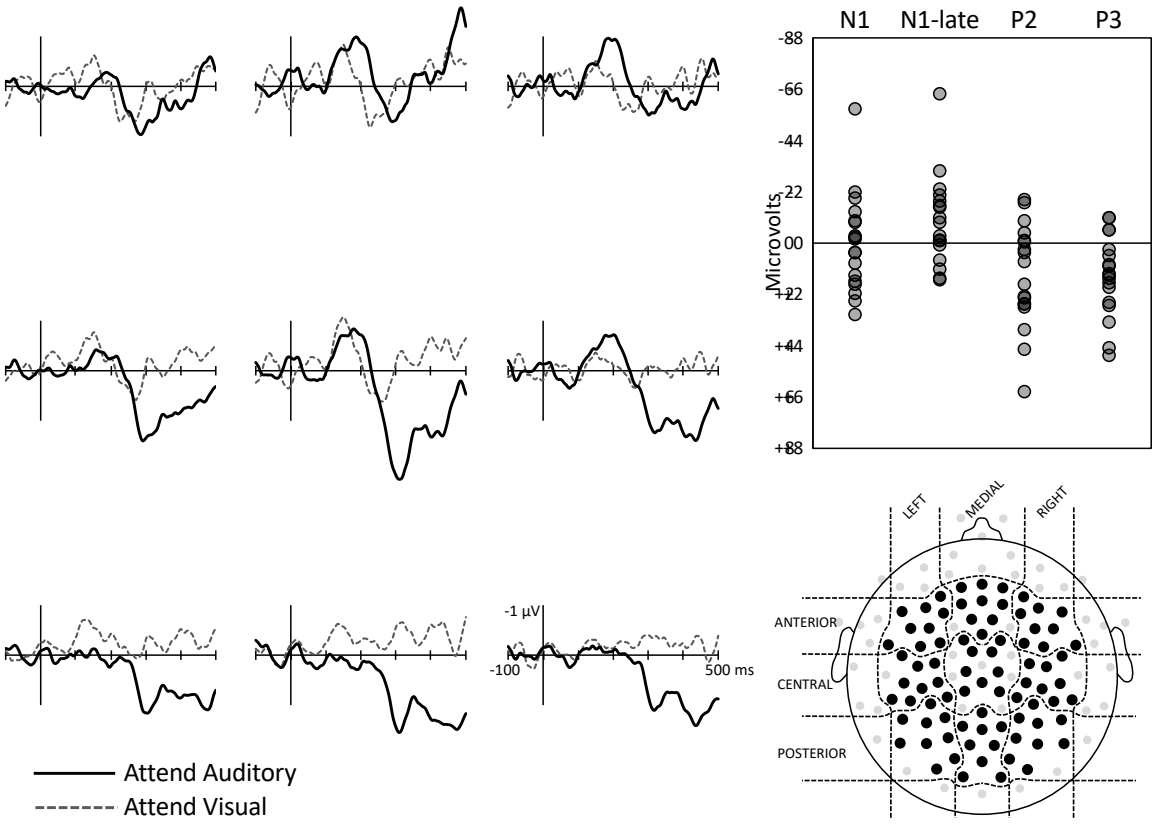

Figure 11: Grand average ERP difference waves representing the spatial condition effects (F-RF minus F-F) for targets presented at SNR_{SRM} in the Attend Auditory and Attend Visual conditions in Study 2 (*N* = 18). Waveforms are averaged across the nine electrodes in a scalp region indicated in the head map and time locked to the earliest abrupt increase in target amplitude (time = 0 ms). A 30-Hz zero-phase low-pass filter was applied to the data shown, but was not part of analysis. In the upper right, the difference in mean amplitude between the difference waves (Attend Auditory minus Attend Visual) for each individual subject is shown for the N1 (130-180 ms), N1-late (200-230 ms), P2 (230-330 ms), and P3 (330-500 ms) time windows.

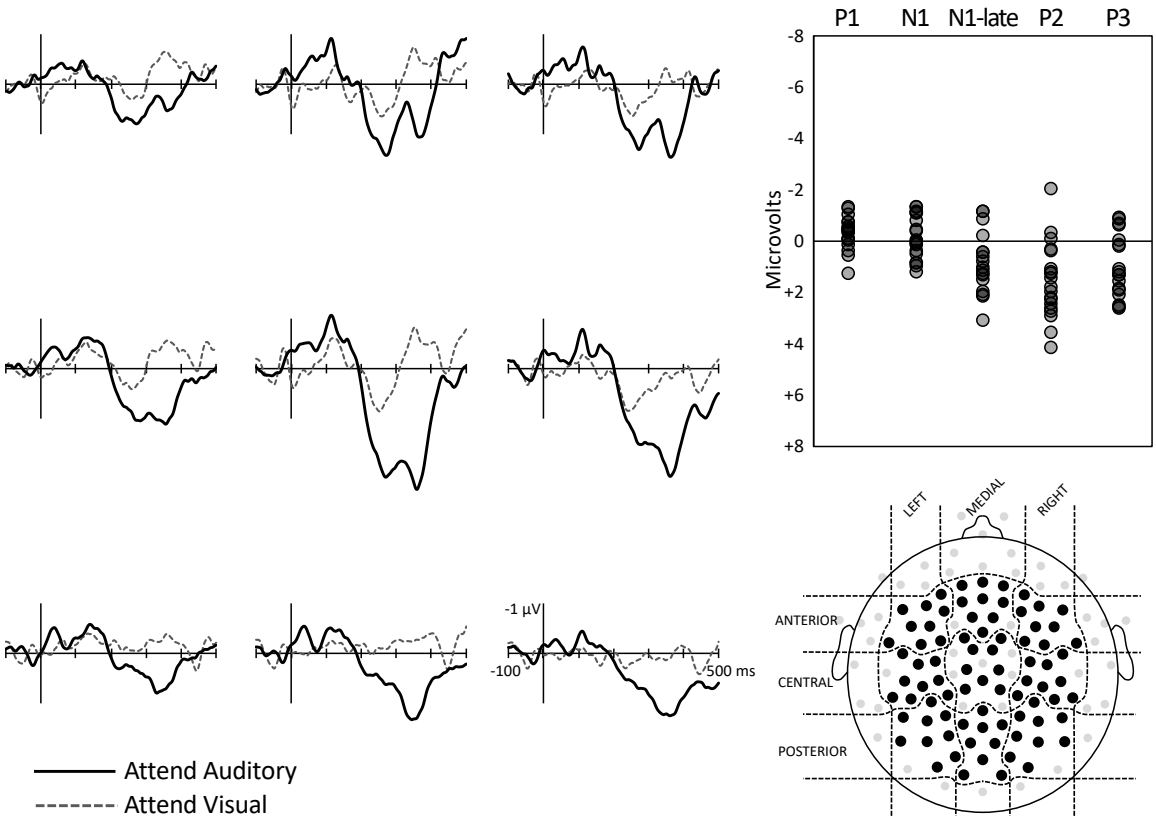

Figure 12: Grand average ERP difference waves representing the spatial condition effects (F-RF minus F-F) for targets presented at SNR_{HIGH} in the Attend Auditory and Attend Visual conditions in Study 2 (*N* = 18). Waveforms are averaged across the nine electrodes in a scalp region indicated in the head map and time locked to the earliest abrupt increase in target amplitude (time = 0 ms). A 30-Hz zero-phase low-pass filter was applied to the data shown, but was not part of analysis. In the upper right, the difference in mean amplitude between the difference waves (Attend Auditory minus Attend Visual) for each individual subject is shown for the P1 (20-60 ms), N1 (130-180 ms), N1-late (200-230 ms), P2 (230-330 ms), and P3 (330-500 ms) time windows.

REFERENCES

Alain, C., Arnott, S. R., & Picton, T. W. (2001). Bottom–up and top–down influences on auditory scene analysis: Evidence from event-related brain potentials. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(5), 1072–1089. https://doi.org/10.1037//TO96-1523.27 j.1072

Alain, C., & Izenberg, A. (2003). Effects of attentional load on auditory scene analysis. *Journal of Cognitive Neuroscience*, *15*(7), 1063–1073. http://www.mitpressjournals.org/doi/abs/10.1162/089892903770007443

Alain, C., Schuler, B. M., & McDonald, K. L. (2002). Neural activity associated with distinguishing concurrent auditory objects. *The Journal of the Acoustical Society of America*, *111*(2), 990–995. https://doi.org/10.1121/1.1434942

Allen, P., & Wightman, F. L. (1995). Effects of Signal and Masker Uncertainty on Children's Detection. *Journal of Speech, Language, and Hearing Research*, *38*(2), 503–511. https://doi.org/10.1044/jshr.3802.503

Arbogast, T. L., Mason, C. R., & Kidd, G., Jr. (2005). The effect of spatial separation on informational masking of speech in normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *117*(4), 2169–2180. https://doi.org/10.1121/1.1861598

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, *39*(3), 445–459. https://doi.org/10.3758/BF03193014

Başkent, D., & Gaudrain, E. (2016). Musician advantage for speech-on-speech perception. *The Journal of the Acoustical Society of America*, *139*(3), EL51–EL56. https://doi.org/10.1121/1.4942628

Bradlow, A. R., & Alexander, J. A. (2007). Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *The Journal of the Acoustical Society of America*, *121*(4), 2339–2349. https://doi.org/10.1121/1.2642103

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

Broadbent, D. E. (1958). *Perception and communication.* Pergamon Press. https://doi.org/10.1037/10037-000

Bronkhorst, A. W. (2000). The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions. *Acta Acustica United with Acustica*, *86*(1), 117–128.

Bronkhorst, A. W. (2015). The cocktail-party problem revisited: Early processing and selection of multi-talker speech. *Attention, Perception, & Psychophysics*, *77*(5), 1465–1487. https://doi.org/10.3758/s13414-015-0882-9

Brungart, D. S. (2005). Informational and Energetic Masking Effects in Multitalker Speech Perception. In P. Divenyi (Ed.), *Speech Separation by Humans and Machines* (pp. 261–267). Springer US. http://dx.doi.org/10.1007/0-387-22794-6_17

Brungart, D. S., & Simpson, B. D. (2007). Cocktail party listening in a dynamic multitalker environment. *Perception & Psychophysics*, *69*(1), 79–91. https://doi.org/10.3758/BF03194455

Brungart, D. S., Simpson, B. D., Ericson, M. A., & Scott, K. R. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. *The Journal of the Acoustical Society of America*, *110*(5), 2527–2538. https://doi.org/10.1121/1.1408946

Brungart, D. S., Simpson, B. D., & Freyman, R. L. (2005). Precedence-based speech segregation in a virtual auditory environment. *The Journal of the Acoustical Society of America*, *118*(5), 3241–3251. https://doi.org/10.1121/1.2082557

Bruns, P. (2019). The Ventriloquist Illusion as a Tool to Study Multisensory Processing: An Update. *Frontiers in Integrative Neuroscience*, *13*, 1–8. https://doi.org/10.3389/fnint.2019.00051

Carhart, R., Tillman, T. W., & Greetis, E. S. (1969). Perceptual masking in multiple sound backgrounds. *The Journal of the Acoustical Society of America*, *45*(3), 694–703. https://doi.org/10.1121/1.1911445

Carlyon, R. P., Cusack, R., Foxton, J. M., & Robertson, I. H. (2001). Effects of attention and unilateral neglect on auditory stream segregation. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(1), 115–127. https://doi.org/10.1037//0096-1523.27.1.115

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the Acoustical Society of America*, *25*(5), 975–979. https://doi.org/10.1121/1.1907229

Crowley, K. E., & Colrain, I. M. (2004). A review of the evidence for P2 being an independent component process: Age, sleep and modality. *Clinical Neurophysiology*, *115*(4), 732–744. https://doi.org/10.1016/j.clinph.2003.11.021

Culling, J. F., & Summerfield, Q. (1995). Perceptual separation of concurrent speech sounds: Absence of across-frequency grouping by common interaural delay. *The Journal of the Acoustical Society of America*, *98*(2), 785–797. https://doi.org/10.1121/1.413571

Darwin, C. J., Brungart, D. S., & Simpson, B. D. (2003). Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers. *The Journal of the Acoustical Society of America*, *114*(5), 2913–2922. https://doi.org/10.1121/1.1616924

Darwin, C. J., & Hukin, R. W. (2000). Effectiveness of spatial cues, prosody, and talker characteristics in selective attention. *The Journal of the Acoustical Society of America*, *107*(2), 970–977. https://doi.org/10.1121/1.428278

Durlach, N. I., Mason, C. R., Kidd, G., Jr., Arbogast, T. L., Colburn, H. S., & Shinn-Cunningham, B. G. (2003). Note on informational masking (L). *The Journal of the Acoustical Society of America*, *113*(6), 2984–2987. https://doi.org/10.1121/1.1570435

Dyson, B. J., Alain, C., & He, Y. (2005). Effects of visual attentional load on low-level auditory scene analysis. *Cognitive, Affective & Behavioral Neuroscience*, *5*(3), 319–338. https://doi.org/10.3758/CABN.5.3.319

El Boghdady, N., Gaudrain, E., & Başkent, D. (2019). Does good perception of vocal characteristics relate to better speech-on-speech intelligibility for cochlear implant users? *The Journal of the Acoustical Society of America*, *145*(1), 417–439. https://doi.org/10.1121/1.5087693

Ericson, M. A., Brungart, D. S., & Simpson, B. D. (2004). Factors That Influence Intelligibility in Multitalker Speech Displays. *The International Journal of Aviation Psychology*, *14*(3), 313–334. https://doi.org/10.1207/s15327108ijap1403_6

Feng, T., Chen, Q., & Xiao, Z. (2018). Age-Related Differences in the Effects of Masker Cuing on Releasing Chinese Speech From Informational Masking. *Frontiers in Psychology*, *9*, 1–16. https://doi.org/10.3389/fpsyg.2018.01922

Festen, J. M., & Plomp, R. (1990). Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *The Journal of the Acoustical Society of America*, *88*(4), 1725–1736. https://doi.org/10.1121/1.400247

Fletcher, H. (1940). Auditory Patterns. *Reviews of Modern Physics*, *12*(1), 47–65. https://doi.org/10.1103/RevModPhys.12.47

Freyman, R. L., Balakrishnan, U., & Helfer, K. S. (2001). Spatial release from informational masking in speech recognition. *The Journal of the Acoustical Society of America*, *109*(5), 2112–2122. https://doi.org/10.1121/1.1354984

Freyman, R. L., Balakrishnan, U., & Helfer, K. S. (2004). Effect of number of masking talkers and auditory priming on informational masking in speech recognition. *The Journal of the Acoustical Society of America*, *115*(5), 2246–2256. https://doi.org/10.1121/1.1689343

Freyman, R. L., Balakrishnan, U., & Helfer, K. S. (2008). Spatial release from masking with noise-vocoded speech. *The Journal of the Acoustical Society of America*, *124*(3), 1627–1637. https://doi.org/10.1121/1.2951964

Freyman, R. L., Helfer, K. S., & Balakrishnan, U. (2005). Spatial and Spectral Factors in Release from Informational Masking in Speech Recognition. *Acta Acustica United with Acustica*, *91*(3), 537–545.

Freyman, R. L., Helfer, K. S., & Balakrishnan, U. (2007). Variability and uncertainty in masking by competing speech. *The Journal of the Acoustical Society of America*, *121*(2), 1040–1046. https://doi.org/10.1121/1.2427117

Freyman, R. L., Helfer, K. S., McCall, D. D., & Clifton, R. K. (1999). The role of perceived spatial separation in the unmasking of speech. *The Journal of the Acoustical Society of America*, *106*(6), 3578–3588. https://doi.org/10.1121/1.428211

Gilmore, C. S., Clementz, B. A., & Berg, P. (2009). Hemispheric differences in auditory oddball responses during monaural versus binaural stimulation. *International Journal of Psychophysiology*, *73*(3), 326–333. https://doi.org/10.1016/j.ijpsycho.2009.05.005

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. John Wiley and Sons, Inc.

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. https://doi.org/10.3758/BF03203619

Hautus, M. J., & Johnson, B. W. (2005). Object-related brain potentials associated with the perceptual segregation of a dichotically embedded pitch. *The Journal of the Acoustical Society of America*, *117*(1), 275–280. https://doi.org/10.1121/1.1828499

Hillyard, S. A., Hink, R. F., Schwent, V. L., & Picton, T. W. (1973). Electrical signs of selective attention in the human brain. *Science*, *182*(4108), 177–180. https://doi.org/10.2307/1736100

Huang, Y., Xu, L., Wu, X., & Li, L. (2010). The effect of voice cuing on releasing speech from informational masking disappears in older adults. *Ear and Hearing*, *31*(4), 579–583. https://doi.org/10.1097/AUD.0b013e3181db6dc2

Ihlefeld, A., & Shinn-Cunningham, B. (2008a). Disentangling the effects of spatial cues on selection and formation of auditory objects. *The Journal of the Acoustical Society of America*, *124*(4), 2224–2235. https://doi.org/10.1121/1.2973185

Ihlefeld, A., & Shinn-Cunningham, B. (2008b). Spatial release from energetic and informational masking in a selective speech identification task. *The Journal of the Acoustical Society of America*, *123*(6), 4369–4379. https://doi.org/10.1121/1.2904826

Kidd, G., Jr., Arbogast, T. L., Mason, C. R., & Gallun, F. J. (2005). The advantage of knowing where to listen. *The Journal of the Acoustical Society of America*, *118*(6), 3804–3815. https://doi.org/10.1121/1.2109187

Kidd, G., Jr., Mason, C. R., Brughera, A., & Hartmann, W. M. (2005). The role of reverberation in release from masking due to spatial separation of sources for speech identification. *Acta Acustica United with Acustica*, *91*(3), 526–536.

Kidd, G., Jr., Mason, C. R., Deliwala, P. S., Woods, W. S., & Colburn, H. S. (1994). Reducing informational masking by sound segregation. *The Journal of the Acoustical Society of America*, *95*(6), 3475–3480. https://doi.org/10.1121/1.410023

Kidd, G., Jr., Mason, C. R., Richards, V. M., Gallun, F. J., & Durlach, N. I. (2008). Informational masking. In W. A. Yost, A. N. Popper, & R. R. Fay (Eds.), *Auditory perception of sound sources* (pp. 143–189). Springer Science+Business Media, LLC.

Kiesel, A., Miller, J., Jolicœur, P., & Brisson, B. (2008). Measurement of ERP latency differences: A comparison of single-participant and jackknife-based scoring methods. *Psychophysiology*, *45*(2), 250–274. https://doi.org/10.1111/j.1469-8986.2007.00618.x

Kraus, N., & Nicol, T. (2008). Auditory evoked potentials. In M. D. Binder, N. Hirokawa, & U. Windhorst (Eds.), *Encyclopedia of Neuroscience* (pp. 214–218). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-29678-2_433

Kutas, M., & Federmeier, K. D. (2011). Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP). *Annual Review of Psychology*, *62*(1), 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical Society of America*, *49*(2B), 467–477. https://doi.org/10.1121/1.1912375

Licklider, J. C. R. (1948). The influence of interaural phase relations upon the masking of speech by white noise. *The Journal of the Acoustical Society of America*, *20*(2), 150–159. https://doi.org/10.1121/1.1906358

Lightfoot, G. (2016). Summary of the N1-P2 cortical auditory evoked potential to estimate the auditory threshold in adults. *Seminars in Hearing*, *37*(01), 1–8. https://doi.org/10.1055/s-0035-1570334

Litovsky, R. Y., Colburn, H. S., Yost, W. A., & Guzman, S. J. (1999). The precedence effect. *The Journal of the Acoustical Society of America*, *106*, 1633–1654. https://doi.org/10.1121/1.427914

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience*, *8*, 1–14. https://doi.org/10.3389/fnhum.2014.00213

Luck, S. J. (2014). *An introduction to the event-related potential technique* (2nd ed.). The MIT Press.

Luck, S. J., & Kappenman, E. S. (2011). The Oxford handbook of event-related potential components. In S. J. Luck & E. S. Kappenman (Eds.), *ERP components and selective attention* (pp. 295–327). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195374148.013.0144

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. https://doi.org/10.1121/1.408204

Lutfi, R. A., Gilbertson, L., Heo, I., Chang, A.-C., & Stamas, J. (2013). The information-divergence hypothesis of informational masking. *The Journal of the Acoustical Society of America*, *134*(3), 2160–2170. https://doi.org/10.1121/1.4817875

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

Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes*, *27*(7–8), 953–978. https://doi.org/10.1080/01690965.2012.705006

Miller, G. A. (1947). The masking of speech. *Psychological Bulletin*, *44*(2), 105–129. pdh. https://doi.org/10.1037/h0055960

Morse-Fortier, C., Parrish, M. M., Baran, J. A., & Freyman, R. L. (2017). The effects of musical training on speech detection in the presence of informational and energetic masking. *Trends in Hearing*, *21*, 1–12. https://doi.org/10.1177/2331216517739427

Newman, R. S., & Evers, S. (2007). The effect of talker familiarity on stream segregation. *Journal of Phonetics*, *35*(1), 85–103. https://doi.org/10.1016/j.wocn.2005.10.004

Oxenham, A. J., Fligor, B. J., Mason, C. R., & Kidd, G. (2003). Informational masking and musical training. *The Journal of the Acoustical Society of America*, *114*(3), 1543–1549. https://doi.org/10.1121/1.1598197

Pashler, H. E. (1999). *The psychology of attention.* The MIT Press.

Picton, T. (2013). Hearing in time: Evoked potential studies of temporal processing. *Ear and Hearing*, *34*(4), 385–401. https://doi.org/10.1097/AUD.0b013e31827ada02

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*(10), 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019

Qin, M. K., & Oxenham, A. J. (2003). Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers. *The Journal of the Acoustical Society of America*, *114*(1), 446–454. https://doi.org/10.1121/1.1579009

Rakerd, B., Aaronson, N. L., & Hartmann, W. M. (2006). Release from speech-on-speech masking by adding a delayed masker at a different location. *The Journal of the Acoustical Society of America*, *119*(3), 1597–1605. https://doi.org/10.1121/1.2161438

Richards, V. M., Huang, R., & Kidd, G. (2004). Masker-first advantage for cues in informational masking. *The Journal of the Acoustical Society of America*, *116*(4), 2278–2288. https://doi.org/10.1121/1.1784433

Richards, V. M., & Neff, D. L. (2003). Cuing effects for informational masking. *The Journal of the Acoustical Society of America*, *115*(1), 289–300. https://doi.org/10.1121/1.1631942

Shaw, E. A. G. (1974). Transformation of sound pressure level from the free field to the eardrum in the horizontal plane. *The Journal of the Acoustical Society of America*, *56*(6), 1848–1861. https://doi.org/10.1121/1.1903522

Singh, G., Pichora-Fuller, M. K., & Schneider, B. A. (2008). The effect of age on auditory spatial attention in conditions of real and simulated spatial separation. *The Journal of the Acoustical Society of America*, *124*(2), 1294–1305. https://doi.org/10.1121/1.2949399

Smulders, F. T. Y. (2010). Simplifying jackknifing of ERPs and getting more out of it: Retrieving estimates of participants' latencies. *Psychophysiology*, *47*(2), 387–392. https://doi.org/10.1111/j.1469-8986.2009.00934.x

Snyder, J. S., Alain, C., & Picton, T. W. (2006). Effects of attention on neuroelectric correlates of auditory stream segregation. *Journal of Cognitive Neuroscience*, *18*(1), 1–13. https://doi.org/10.1162/089892906775250021

Treisman, A. (1964). Selective attention in man. *British Medical Bulletin*, *20*(1), 12–16. https://doi.org/10.1093/oxfordjournals.bmb.a070274

Tremblay, K., Ross, B., Inoue, K., McClannahan, K., & Collet, G. (2014). Is the auditory evoked P2 response a biomarker of learning? *Frontiers in Systems Neuroscience*, *8*, 1–13. https://doi.org/10.3389/fnsys.2014.00028

Vestergaard, M. D., Fyson, N. R. C., & Patterson, R. D. (2009). The interaction of vocal characteristics and audibility in the recognition of concurrent syllables. *The Journal of the Acoustical Society of America*, *125*(2), 1114–1124. https://doi.org/10.1121/1.3050321

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

Watson, C. S., Kelly, W. J., & Wroton, H. W. (1976). Factors in the discrimination of tonal patterns. II. Selective attention and learning under various levels of stimulus uncertainty. *The Journal of the Acoustical Society of America*, *60*(5), 1176–1186. https://doi.org/10.1121/1.381220

Watson, C. S., Wroton, H. W., Kelly, W. J., & Benbassat, C. A. (1975). Factors in the discrimination of tonal patterns. I. Component frequency, temporal position, and silent intervals. *The Journal of the Acoustical Society of America*, *57*(5), 1175–1185. https://doi.org/10.1121/1.380576

Wightman, F. L., Callahan, M. R., Lutfi, R. A., Kistler, D. J., & Oh, E. (2003). Children's detection of pure-tone signals: Informational masking with contralateral maskers. *The Journal of the Acoustical Society of America*, *113*(6), 3297–3305. https://doi.org/10.1121/1.1570443

Wightman, F. L., & Kistler, D. J. (2005). Informational masking of speech in children: Effects of ipsilateral and contralateral distracters. *The Journal of the Acoustical Society of America*, *118*(5), 3164–3176. https://doi.org/10.1121/1.2082567

Woldorff, M. G., & Hillyard, S. A. (1991). Modulation of early auditory processing during selective listening to rapidly presented tones. *Electroencephalography and Clinical Neurophysiology*, *79*(3), 170–191. https://doi.org/10.1016/0013-4694(91)90136-R

Wolpaw, J. R., & Penry, J. K. (1977). Hemispheric differences in the auditory evoked response. *Electroencephalography and Clinical Neurophysiology*, *43*(1), 99–102. https://doi.org/10.1016/0013-4694(77)90200-0

Woods, D. L. (1990). The physiological basis of selective attention: Implications of event-related potential studies. In J. W. Rohrbaugh, R. Parasuraman, & R. Jr. Johnson (Eds.), *Event-related brain potentials: Basic issues and applications.* (1990-99035-013; pp. 178–209). Oxford University Press.

Wu, M., Li, H., Gao, Y., Lei, M., Teng, X., Wu, X., & Li, L. (2012). Adding irrelevant information to the content prime reduces the prime-induced unmasking effect on speech recognition. *Hearing Research*, *283*(1), 136–143. https://doi.org/10.1016/j.heares.2011.11.001

Wu, M., Li, H., Hong, Z., Xian, X., Li, J., Wu, X., & Li, L. (2012). Effects of aging on the ability to benefit from prior knowledge of message content in masked speech recognition. *Speech Communication*, *54*(4), 529–542. https://doi.org/10.1016/j.specom.2011.11.003

Wu, X., Wang, C., Chen, J., Qu, H., Li, W., Wu, Y., Schneider, B. A., & Li, L. (2005). The effect of perceived spatial separation on informational masking of Chinese speech. *Hearing Research*, *199*(1–2), 1–10. https://doi.org/10.1016/j.heares.2004.03.010

Yang, Z., Chen, J., Huang, Q., Wu, X., Wu, Y., Schneider, B. A., & Li, L. (2007). The effect of voice cuing on releasing Chinese speech from informational masking. *Speech Communication*, *49*(12), 892–904. https://doi.org/10.1016/j.specom.2007.05.005

Yonan, C. A., & Sommers, M. S. (2000). The Effects of Talker Familiarity on Spoken Word Identification in Younger and Older Listeners. *Psychology and Aging*, *15*(1), 88–99. https://doi.org/10.1037/0882-7974.15.1.88

Zhang, C., Arnott, S. R., Rabaglia, C., Avivi-Reich, M., Qi, J., Wu, X., Li, L., & Schneider, B. A. (2016). Attentional modulation of informational masking on early cortical representations of speech signals. *Hearing Research*, *331*, 119–130. https://doi.org/10.1016/j.heares.2015.11.002

Zhang, C., Lu, L., Wu, X., & Li, L. (2014). Attentional modulation of the early cortical representation of speech signals in informational or energetic masking. *Brain and Language*, *135*, 85–95. https://doi.org/10.1016/j.bandl.2014.06.002

Zhang, C., Tao, R., & Zhao, H. (2019). Auditory spatial attention modulates the unmasking effect of perceptual separation in a "cocktail party" environment. *Neuropsychologia*, *124*, 108–116. https://doi.org/10.1016/j.neuropsychologia.2019.01.009

Zobel, B. H., Freyman, R. L., & Sanders, L. D. (2015). Attention is critical for spatial auditory object formation. *Attention, Perception, & Psychophysics*, *77*(6), 1998–2010. https://doi.org/10.3758/s13414-015-0907-4

Zobel, B. H., Wagner, A., Sanders, L. D., & Başkent, D. (2019). Spatial release from informational masking declines with age: Evidence from a detection task in a virtual separation paradigm. *The Journal of the Acoustical Society of America*, *146*(1), 548–566. https://doi.org/10.1121/1.5118240

Zurek, P. M. (1993). Binaural advantages and directional effects in speech intelligibility. In G. A. Studebaker & I. Hochberg (Eds.), *Acoustical factors affecting hearing aid performance* (2nd ed., pp. 255–276). Allyn & Bacon.