UNIVERSITY OF CALIFORNIA,
IRVINE

Investigating level dominance through psychophysics and electrophysiology

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Cognitive Sciences

by

Kenta Watanabe

Dissertation Committee:
Professor Virginia Richards, Chair
Professor Ramesh Srinivasan
Professor John Middlebrooks

2024

DEDICATION

This dissertation is dedicated to my family, friends and mentors, and importantly, myself.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

I would like to express my gratitude and appreciation to my advisor Virginia Richards for her support and mentorship. She provided me with a solid foundation in hearing science, allowing me to continue pursuing research in this field. I would also like to express my appreciation to my committee members, Ramesh Srinivasan and John Middlebrooks who have provided time, critical insight, and guidance in this work.

I extend a special thanks to additional members of my advancement committee, Bruce Berg and Raju Metherate.

I would also like to acknowledge and thank the UCI Center for Hearing Research for their training and funding support for this dissertation under the NIH Training Grant T32-DC010775.

Financial support was provided by the University of California, Irvine, laboratory grants/funds, and the Center for Hearing Research.

VITA

Kenta Watanabe

EDUCATION

Doctor of Philosophy in Cognitive Sciences **December 2024**
with an emphasis in Neuroscience
University of California, Irvine *Irvine, CA*

Master of Arts in Psychology **September 2019**
Animal Behavior and Conservation
City University of New York, Hunter College *New York, NY*

Bachelor of Arts in Psychology **March 2017**
Minored in Physics
International Christian University *Tokyo, Japan*

EMPLOYMENT

Teaching Assistant **Oct 2019–June 2020**
University of California, Irvine *Irvine, CA*
Courses: Perception Lab, Intro to Psych, Adolescent Psych

Research Assistant **March 2017–July 2017**
International Christian University *Tokyo, Japan*

FUNDING AND AWARDS

T32 NIH-NRSA Predoctoral Training Fellowship **2020–2024**
Center for Hearing Research, UC Irvine

Doctoral Student Conference Travel Funds **2023**
The School of Social Sciences, UC Irvine

Graduate Student Research Fellowship **2019–2020**
Graduate Division, UC Irvine

Minority Serving Institution Enhancement (MSIE) Award **2019**
Graduate Division, UC Irvine

Animal Behavior and Conservation (ABC) Research Grants **2017**
Psychology Department, CUNY Hunter College

MENTORSHIP

Undergraduate research supervision:

Temporal and spectral model for sound discrimination **2021–2022**
Summer Undergraduate Research Program (UROP)

JOURNAL ARTICLES

Effects of sounds segregation cues on multi-sound intensity discrimination **2023**
JASA Express Letters, Selected as Editor's Pick

CONFERENCE POSTERS

Attention allocation and level dominance: insights from an EEG study **November 2023**
Society for Neuroscience 2023

Loudness dominance and perceptual dissimilarity **February 2023**
46th annual midwinter meeting of the Association for Research in Otolaryngology

Is there synergy between song learning and vocal stimuli discrimination training? **November 2018**
Society for Neuroscience 2018

Investigating the effect of difficulty levels of inference in second language processing on the frontal region activation **July 2017**
20th annual meeting of Japan Optical Functional Brain Imaging Society in Tokyo, Japan

ABSTRACT OF THE DISSERTATION

Investigating level dominance through psychophysics and electrophysiology

By

Kenta Watanabe

Doctor of Philosophy in Cognitive Sciences

University of California, Irvine, 2024

Professor Virginia Richards, Chair

This dissertation explores an auditory phenomenon known as *level dominance*, where listeners assign a significantly higher perceptual weight to louder sounds than softer ones in a sequence. The first two chapters present psychophysical studies, while the final chapter involves electrophysiological (EEG) investigation. Chapter 1 examines the impact of changes in sound qualities on level dominance through a multi-sound intensity discrimination task. The results indicate that changes in pitch or timbre do not reduce level dominance; however, manipulations that reduce frequency overlap between sounds lead to a notable release from level dominance. Chapter 2 investigates the relationship between level dominance and stream segregation tasks using various sound manipulations. Both individual and groupwise correlational analyses reveal a significant association between the two tasks, suggesting the involvement of common underlying mechanisms. Chapter 3 shifts focus to the electrophysiological domain, exploring how the relative levels of sequential sounds influence selective attention. This chapter employs EEG to uncover the neural correlates of selective attention in the context of level dominance. Behavioral results showed that accuracy was high only when listeners attended to high-level targets. EEG data indicated that listeners' attention was directed toward louder components even when they were instructed to attend to the softer components.

Chapter 1

Effects of sound segregation cues on multi-sound intensity discrimination

(This chapter is written based on Watanabe, K., Srinivasan, R., & Richards, V. M. (2023). Effects of sound segregation cues on multi-sound intensity discrimination. JASA Express Letters, 3(12)., with the permission of AIP Publishing)

1.1 Introduction

For a temporal series of short sounds, the more intense sounds may be overweighted when integrating information across the series of sounds. This general effect is referred to as level dominance (Berg, 1990; Ponsot et al., 2013; Oberfeld et al., 2013; Richards et al., 2013). In a frequency sample-discrimination task, Berg (1990) estimated perceptual weights for a series of tone pips drawn from one or another frequency distribution. The tones were presented sequentially and varied in level: high-low-high-low-high-low-high. Listeners' responses were dominated by the frequencies associated with the higher-level tones, even when the lower-level tones were more informative and reliable (i.e., drawn from a distribution with smaller frequency variation).

Understanding this phenomenon involves exploring both peripheral mechanisms, those occurring at the level of the basilar membrane, and central or cognitive mechanisms, involving higher-level auditory structures such as superior olivary nuclei, inferior colliculus, or auditory cortex. At the peripheral level, the presence of a sound degrades the sensitivity to a subsequent sound (forward masking), which is largely due to mechanical and neural processes in the cochlea. For instance, Zeng et al. (1991) proposed a model based on adaptation in the auditory nerve and showed that the elevation of intensity-difference limens (DL) at mid-level intensities, known as the 'mid-level hump', is caused by the adaptation of low-spontaneous rate (SR) auditory nerve fibers, which recover more slowly than the high-SR population (Relkin and Doucet, 1991).

Whereas peripheral mechanisms are expected to contribute to level dominance (e.g., forward masking), there is evidence that higher-level, central/cognitive processes also contribute. For example, Oberfeld (2008) demonstrated that there exists a 'mid-difference' hump phenomenon, where significant DL elevations occur at the intermediate masker-standard difference, and this is incompatible with the recovery-rate model (Zeng et al., 1991). In a multitone intensity discrimination task, Oberfeld et al. (2013) found that level dominance is associated with centrally generated noise for some listeners. Additionally, sequential grouping of the maskers and targets into separate auditory objects (Oberfeld and Stahn, 2012) and perceived lateralization due to the interaural time difference (Oberfeld et al., 2012) led to a release from masking. For frequency sample discrimination tasks, Turner and Berg (2007) showed that level dominance is observed with inter-sound intervals well beyond 300 ms, which is unlikely to be attributable solely to forward or backward masking.

Lutfi and Jesteadt (2006) and Tan and Berg (2018) examined the effect of the composition/presentation of the higher- and lower-level sounds. Lutfi and Jesteadt (2006) explored level dominance using a multitone intensity discrimination task. A series of sounds were presented such that their intensities varied high-low-high-low-high. The intensity increment applied to the lower-level tones was larger than that for the higher-level tones. Nonetheless, listeners overweight the information in the more intense sounds. However, when the higher-level tone pips were replaced by broadband noise, listeners more efficiently used the information from the lower-level tones. Tan and Berg (2018) used a sample discrimination task and tested a condition where less intense, more informative tones and more intense, less informative tones alternate between ears. The results showed that dichotic presentations also release level dominance. In both instances, one might suggest that the salience of the higher-level sounds can be overcome by introducing differences in the quality of the higher- versus lower-level sounds, allowing the listener to attend to the less salient lower-level sounds. Potentially, manipulations known to promote auditory stream segregation may also be effective.

Auditory stream segregation is a fundamental process by which the auditory system organizes sequential sounds into two different streams (Bregman et al., 1990), allowing individuals to distinguish between different sound sources in complex acoustic environments. There are two primary mechanisms facilitating this process: compulsory (also referred to as obligatory or primitive) stream segregation and voluntary stream segregation.

Compulsory stream segregation is considered an automatic, bottom-up process driven by the acoustic properties of sound stimuli (Moore and Gockel, 2012). This type of segregation occurs without conscious effort and provides a rapid parsing of the auditory scene, as opposed to voluntary stream segregation, which requires conscious effort to organize auditory input based on relevance or interest (Shinn-Cunningham, 2008). Some of the key acoustic cues influencing compulsory segregation include frequency difference (e.g., van Noorden, 1975), pitch difference (e.g., Vliegen and Oxenham, 1999), or timbre difference (e.g., Iverson, 1995). Differences in frequency content between sounds are among the most potent cues for stream segregation (Moore and Gockel, 2012). This is reflected in the peripheral channeling model (Hartmann and Johnson, 1991; Beauvois and Meddis, 1996) which states that stream segregation occurs when the energy of sequential sounds falls in distinct auditory channels. However, frequency difference is not required for stream segregation. The studies indicated that compulsory stream segregation can occur even when differences in pitch (Summers et al., 2010) or timbre (Iverson, 1995; Cusack and Roberts, 2000) are not associated with differences in frequency regions (e.g., two sounds occupy wholly overlapping frequency regions), providing evidence against a strict peripheral channel model. This study aimed to examine whether these sound properties (changes in pitch, timbre, and frequency regions) known to promote compulsory stream segregation likewise reduce level dominance.

The current experiments are somewhat parallel to those of Lutfi and Jesteadt (2006). A series of sounds with alternating levels were used, and the increment in level to be detected was applied to the second and fourth of five sounds. The goal is to evaluate the role of changes in sound "quality" on intensity discrimination. The makeup of the nontarget sounds was altered to systematically influence the difference in quality of the target and nontarget sounds. As a proxy for changes in quality, the makeup of the nontarget sounds varied from the target sounds along the continuum of changes that are expected to promote stream segregation or not. Notably, the target and nontarget sounds within the sequence were perceptibly different. However, they were heard as a sequence, not as two streams due to the relatively long temporal gaps between sounds and the small number of sounds (five) per trial.

In the first experiment, the composition of the higher- and lower-level sounds was varied using manipulations tested by Oh et al. (2022), who systematically evaluated the impact of changes in fundamental frequency and spectral slope on stream segregation (see also Bregman et al., 1990; Vliegen and Oxenham, 1999; Moore and Gockel, 2002). They demonstrated that a greater difference in pitch or timbre between sounds results in a greater amount of stream segregation. Building on this finding, we tested whether the same separation in pitch or timbre leads to a reduction in level dominance as well. The results indicate that neither differences in fundamental frequency nor spectral slope led to changes in intensity discrimination thresholds. In the second experiment, the spectral extent of the target and nontarget sounds differed, providing strong peripheral cues for differences between the two sounds. The results indicate consistent releases from masking for these conditions.

1.2 Methods

1.2.1 Participants

A total of 12 subjects (not including the excluded subjects described later) completed the current experiments, half of whom participated in more than one experiment. Six participated in experiment 1 (aged 18–30 years old, four males), seven participated in experiment 2A (aged 18–29 years old, one male), and six participated in experiment 2B (aged 18–30 years old, three males). The subjects were paid for participation, except for K.W., the author. All subjects had absolute thresholds of 20 dB hearing level (HL) or better for audiometric frequencies between 250 and 8000 Hz.

1.2.2 Stimuli and design

The stimuli were digitally generated using a sampling frequency of 44100 Hz on a personal computer (PC), which also controlled the experimental procedure and data collection through custom-written software (The MathWorks, Inc., Natick, MA). The stimuli were presented diotically via a 24-bit soundcard (Envy 24 PCI audio controller, VIA Technologies, Inc., Taipei, Taiwan), a programmable attenuator and headphone buffer (PA4 and HB6, Tucker-Davis Technologies, Inc., Alachua, FL), and Sennheiser HD 600 headsets (Wennebostel, Germany). Each stimulus presentation was followed by visual feedback as to the correctness of the subject's response. The experiment was conducted in a double-walled, sound-attenuated booth.

A single-interval, yes-no, intensity discrimination procedure was used. Figure 1.1 is a schematic of the stimuli used, plotting stimulus intensity as a function of time. Each stimulus was composed of five 100-ms sounds with 5-ms cosine-squared onset and offset ramps, separated by 300 ms. In experiment 1, the standard stimulus (i.e., no signal present) had sound levels that alternated between 70- and 35-dB sound pressure level (SPL), i.e., 70-35-70-35-70-dB SPL [Fig. 1.1(A)]. This series of levels will be referred to as the low target (LT) stimulus because the 35-dB SPL sounds were the target. In experiment 2, LT [Fig. 1.1(A)] and high target [Fig. 1.1(B)] stimuli were tested. For the HT conditions, sounds had alternating levels of 35-70-35-70-35-dB SPL, and the intensity increment was applied to the 70-dB SPL sounds.

The magnitude of the level increment (as ΔL in dB) was adjusted using a two-down, one-up adaptive procedure (Levitt, 1971). The initial value of ΔL was 3.5 dB, and the step size was 0.5 dB; after four reversals, the step size was reduced to 0.25 dB and the track continued

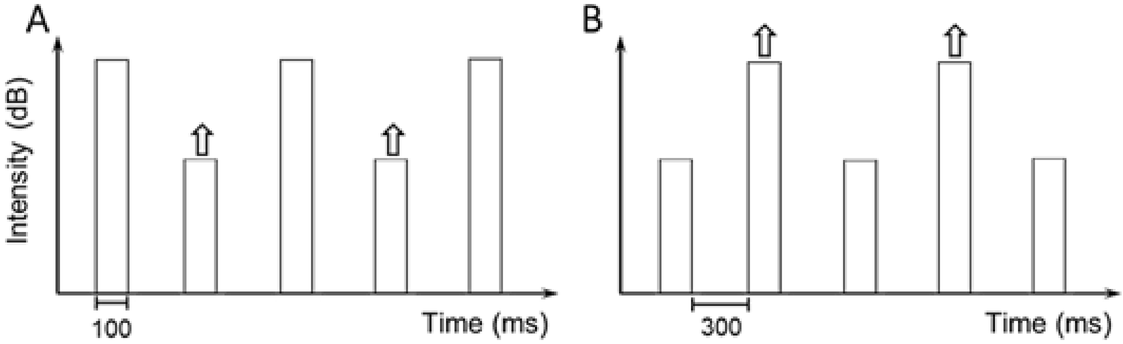

Figure 1.1: Schematic of two stimuli tested. (A) Low target (LT) stimulus, where an increment in level (indicated by arrows) is added to the lower-level target sounds in the presence of three higher-level nontarget sounds. (B) High target (HT) stimulus, where an increment is added to the higher-level target sounds in the presence of three lower-level nontarget sounds.

until an additional eight reversals were completed. The 71% correct threshold is taken as the arithmetic mean of the last eight reversals. The minimum value of ΔL was set to 0 dB to prevent negative values. In the baseline conditions, only the target sounds were present; the first, third, and fifth sounds were replaced with temporal gaps.

1.2.3 Experiment 1

For experiment 1, the stimuli were drawn from Oh et al. (2022). The target sounds (Figure 1.2, top row) were the 1st – 16th harmonics of 250 Hz fundamental frequency (f_o) and the slope of the spectrum that fell at a rate of 1 dB/octave with phases drawn randomly prior to each stimulus presentation. There were no harmonics presented above 4000 Hz.

The characteristics of the nontarget sounds (Figure 1.2, bottom row) varied to form different conditions. In the *same* condition, the target and nontarget sounds had the same fundamental frequency and harmonic numbers. In the *low*, *lower*, and *lowest* conditions, the f_o's of the nontarget sounds were 198, 176, or 136, respectively, with no harmonics presented above 4000 Hz, and the slope of the spectrum fell at 1 dB/octave. In the *steep*, *steeper*, and *steepest*

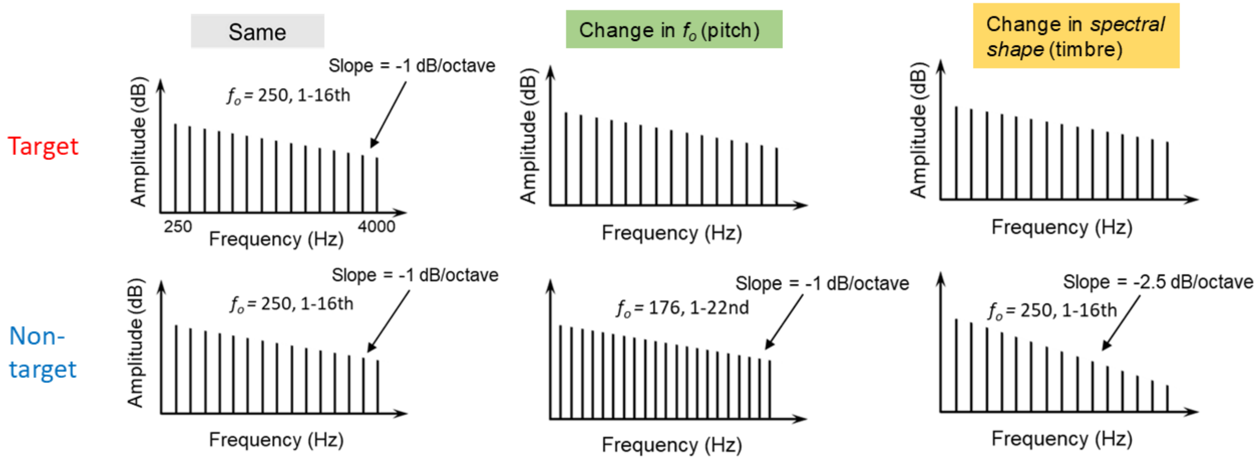

Figure 1.2: Description of sound stimuli and conditions in experiment 1. The target sounds were the 1st – 16th harmonics of $f_o = 250$ Hz with the spectral slope of -1 dB/octave. The characteristics of the nontarget sounds varied to form different conditions: *Same*, change in pitch, and change in timbre.

conditions, nontargets were the 1st–16th harmonics of $f_o = 250$ Hz, and the slope of the spectrum fell at a rate of 1.9, 2.5, or 7 dB/octave, respectively. Among the stimuli, all but the *same* conditions are considered compulsory stream segregation cues. The values for the fundamental frequencies and spectral slopes were chosen from Oh et al. (2022), where the smallest changes did not uniformly lead to stream segregation while the two larger changes did. We tested whether such a pattern would be reflected in the intensity discrimination thresholds.

The protocol for experiment 1 was as follows. Initially, listeners practiced in the baseline condition. The first five thresholds were averaged. If the average was greater than $\Delta L = 2$ dB, another five thresholds were estimated. If the listener's threshold remained above ΔL = 2 dB after 2 hours of practice, the listener was excluded from the study. For experiment 1, two subjects were excluded. Next, listeners ran the baseline condition. After completion, the listeners ran the remaining conditions in random orders. For the initial non-baseline condition encountered by a listener, 13 threshold estimates were collected; otherwise, 10 threshold estimates were collected. The last eight threshold estimates are split into two blocks of four to test for practice effects (one-tailed t-test). If no practice effects were noted, the average of those eight threshold estimates formed the threshold estimate. On four occasions, practice effects were observed. In these cases, an additional five threshold estimates were collected and the last eight were averaged to estimate the threshold.

1.2.4 Experiment 2

Experiment 2 is divided into two parts: experiment 2A and experiment 2B (Figure 1.3). For experiment 2A, the targets were the 1st – 15th harmonics of a 250-Hz fundamental frequency. Three conditions were tested in addition to the baseline and *same* conditions. In the noise condition, the nontarget sounds were broadband Gaussian noise (after Lutfi and Jesteadt, 2006). The noise was not low-pass filtered prior to headphones. In the low and lower conditions, the nontarget sounds were the 1st–15th harmonics of f_o = 198 and 176 Hz, respectively. Note that because the number of harmonics is the same, stimuli with higher fundamental frequencies had wider bandwidths. In experiment 2B, the targets were the third–seventh harmonics of a 250-Hz fundamental. In addition to the baseline condition, in the *off-freq* condition, the nontarget sounds were the 10th – 14th harmonics of a 250-Hz fundamental, i.e., having spectral energy that shares no frequencies (non-overlapping frequency).

For ease of comparison across conditions, we quantified the proportion of shared frequencies between target and nontarget sounds by dividing the bandwidth of whichever sound with the narrower spectrum by that of the sound with the wider spectrum so that the values range from 0 to 1. For the *noise* condition, the bandwidth of Gaussian noise was set at 20000Hz.

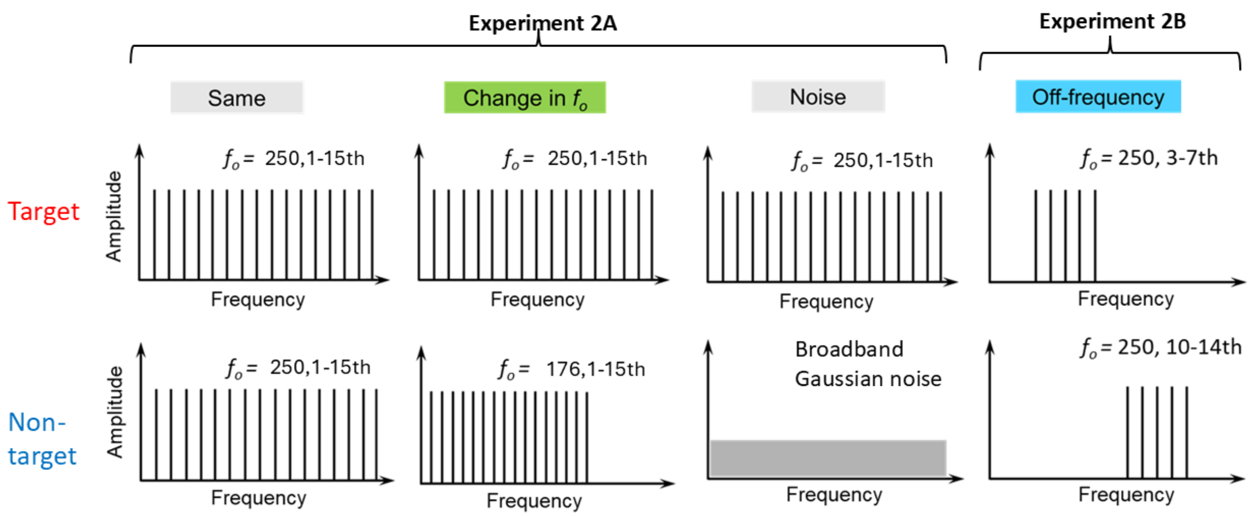

Figure 1.3: Description of sound stimuli and conditions in experiment 2. The target sounds were the harmonic complexes of $f_o = 250$ Hz with the flat spectrum (composed of 1-15th harmonics in experiment 2A, and 3-7th harmonics in experiment 2B). The characteristics of the nontarget sounds varied to form three types of conditions besides *Same*: change in f_o associated with a change in frequency region, *noise*, and *off-frequency*.

Thus, the proportion of shared frequency was computed as follows: *same* (1.0), *low* (0.79), *lower* (0.70), *noise* (0.19), and *off-freq* (0.0).

The protocols for experiments 2A and 2B were as follows. In experiment 2A, three listeners ran the LT conditions and then the HT conditions; for four listeners, the order was the opposite. In experiment 2B, three listeners ran the LT conditions and then the HT conditions; for the other three listeners, the order was the opposite. In other respects, the procedures were parallel to those described for experiment 1. Listeners' practice thresholds were sufficiently low that none were excluded from completing the experiment. On one occasion in experiment 2A and two occasions in experiment 2B, practice effects were observed in the experimental condition and repeated as described above.

The *low* and *lower* conditions of experiments 1 and 2 share the same fundamental frequencies ($f_o = 198$ and 176 Hz, respectively). In experiment 1, the spectral extent (cutoff frequency) is the same for the target and nontargets sounds in the *low* and *lower* conditions. In contrast, in the low and lower conditions of experiment 2, the nontarget sounds have narrower bandwidths than the targets (by 780 and 1110 Hz, respectively). Comparing the *same*, *low* and *lower* conditions across experiments provides an opportunity to evaluate potential cues associated with differences in frequencies per se (experiment 1, the spectral extent is similar, but the harmonic components are not) versus changes in spectral regions associated at higher frequencies (experiment 2, same range of harmonic numbers and different bandwidths).

To the degree that changes in magnitude follow the strength of stream segregation, the predictions for these experiments are as follows. For experiments 1 and 2, the difference in thresholds between the *baseline* and *same* conditions is expected to be the largest. Relative to that value, threshold differences between the *same* and experimental conditions would reflect a release from level dominance. Note that the target sounds are unchanged, meaning any shift in thresholds reflects the properties of the nontarget sounds (parallel to Lutfi and Jesteadt, 2006).

1.3 Results

1.3.1 Experiment 1

Figure 1.4 plots the mean thresholds (LT) for all conditions; *baseline*, *same*, and the three conditions with differences in fundamental frequency and spectral slope (left to right), smallest to largest change, respectively. As expected, thresholds in the baseline condition are the lowest. Adding nontarget sounds (*same*) yields a 2.5 dB increase in threshold relative to the baseline condition. A one-way repeated measures ANOVA was performed to investigate the effect of the differences in fundamentals and spectral slope (all the conditions except

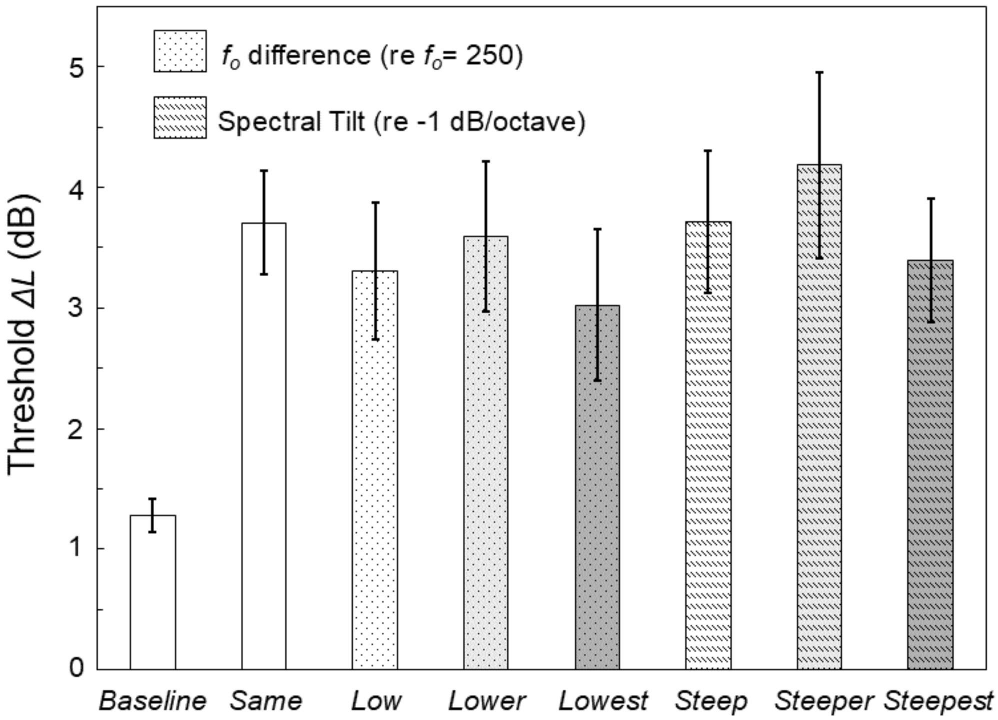

Figure 1.4: Thresholds for experiment 1 as ΔL in dB. In different LT conditions, nontarget sounds are absent (baseline), nontarget and target sounds are the same (*same*), nontarget sounds have different fundamental frequencies (*low*, *lower*, and *lowest* conditions), or different spectral slopes (*steep*, *steeper*, and *steepest* conditions) relative to target sounds. Error bars indicate ± one standard error of the mean. Statistical analysis is described in the text.

baseline) on threshold measurements and no significant main effect was revealed, $F(5,25) = 2.35$, $p = .07$. The thresholds associated with changes in the fundamental frequency and spectral slope of the nontargets do not demonstrate graded releases from level dominance, i.e., there is no consistent reduction in thresholds as the fundamental frequency/spectral slope of the nontarget sounds deviate more and more from the targets. This is counter to the proposed expectations based on primary sound segregation results (Oh et al., 2022).

Potentially, the changes in stimuli tested here do not provide a release from level dominance because the perceptual difference between the target and nontarget sounds is not large enough and, thus, failed to overcome the salience of the higher-level nontargets. To explore this possibility, in experiment 2, target and nontarget sounds differed in terms of the extension of energy in non-overlapping frequency regions.

1.3.2 Experiment 2

Across conditions in experiment 2A, threshold ΔLs for the HT stimulus were within 0.1 dB of each other, including the baseline condition. That is, thresholds with and without low-level nontargets were the same regardless of the characteristics of the low-level nontargets (see also Oberfeld, 2008 for parallel results in forward masking).

Figure 1.5 plots the difference in LT thresholds minus HT thresholds. The left side shows results for experiment 2A, and the right two bars are for experiment 2B. The leftmost five bars are *baseline*, *same*, *low* and *lower* with nontarget fundamental frequencies of $f_o = 198$ and 176 and the 1st–15th harmonics (highest frequency of 2970 and 2640 Hz, respectively), and noise (nontargets were noise bursts). The right side shows results for the baseline (different target sounds than in experiment 2A) and *off-freq* conditions in experiment 2B. The proportion of shared frequencies between target and nontarget sounds is indicated at the bottom of the Figure 1.5.

Because thresholds across the HT conditions were essentially identical, Fig. 1.5 effectively reflects threshold profiles for the LT conditions. For experiment 2A, a one-way repeated measures ANOVA (excluding *baseline*) indicated a significant effect of the condition on threshold difference, $F(3,18) = 5.97$, $p < .01$. Because our primary interest was to determine whether there was a significant release from the *same* condition, we conducted planned comparisons between the *same* condition with each of the three non-*baseline* conditions. One-tailed paired t-test after applying a Bonferroni correction for multiple comparisons (adjusted alpha = 0.0167) showed no significant difference between *same* and *low* ($p = .047$), but the

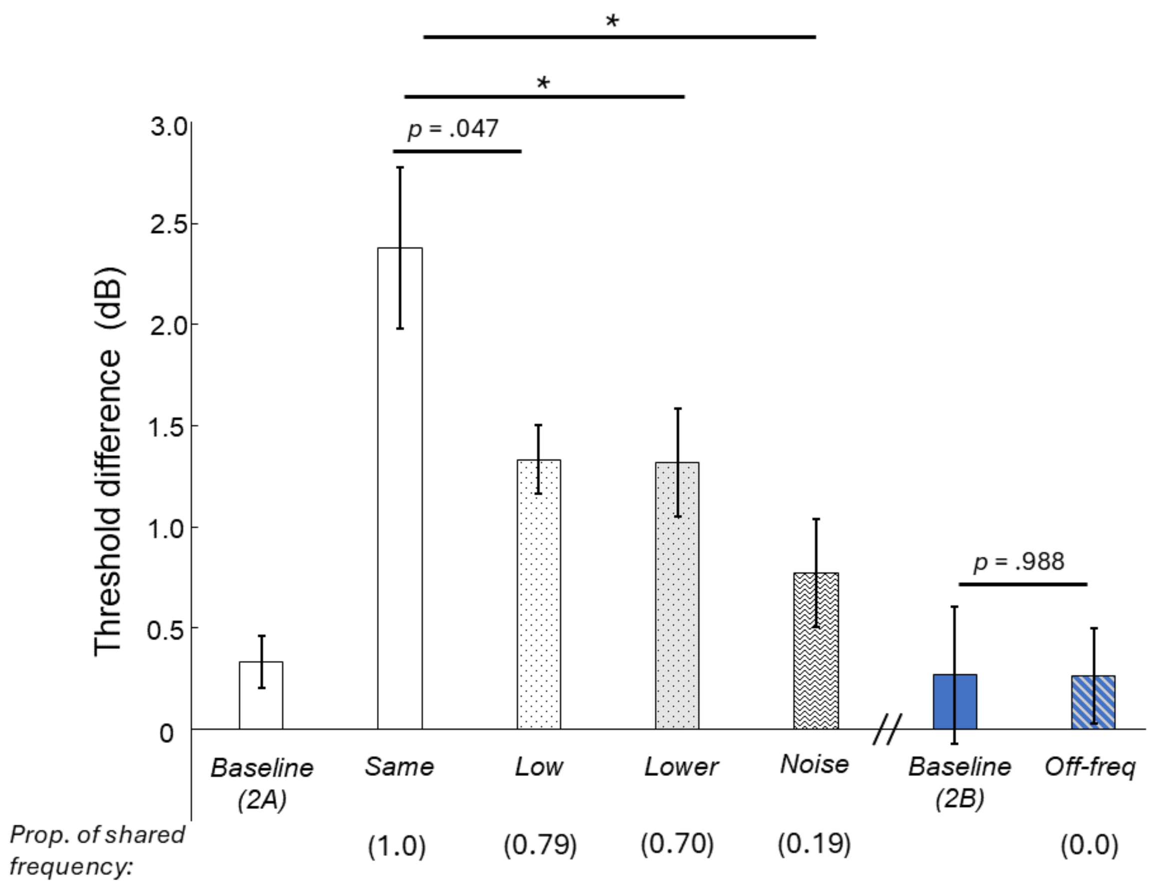

Figure 1.5: Differences in thresholds, LT minus HT. The first five bars are from experiment 2A, and the rightmost two bars are from experiment 2B. Relative to the target sounds, the nontarget sounds are absent (*baseline*); the same (*same*); of different fundamental frequencies but the same harmonic numbers (*low*, *lower*); broadband noise (*noise*); of the same harmonic frequency but different harmonic numbers (*off-freq*). Error bars indicate ± one standard error of the mean. Statistical analyses are described in the text. Significant differences from the *same* condition are indicated by *, while non-significant differences are labeled with their p-value. The values at the bottom indicate the proportion of shared frequency.

difference was significant between *same* and *lower* ($p = .012$) and *same* and *noise* ($p < .01$). Unlike the results of experiment 1, in the current experiment with equal harmonic numbers, changing the fundamental frequencies improved the thresholds. Presumably, this reflects not a change in the fundamental frequency but a shift in the frequency extent for the target relative to the nontarget sounds. For the noise condition, the level dominance was further reduced (as in Lutfi and Jesteadt, 2006). The two rightmost bars in Fig. 1.5 are for experiment 2B in which the target sounds were harmonic complexes with a fundamental frequency of 250 Hz and the 3rd–7th (750-1750 Hz) harmonics. For the *off-freq* condition, the added nontargets shared the 250-Hz fundamental with the targets but were composed of the 10th–14th (2500-3500 Hz) harmonics. A pairwise t-test showed no significant difference between *baseline* and *off-freq* conditions ($p = .988$). When the nontarget sounds and target sounds had distinct regions of activation, a threshold difference was essentially identical to *baseline*, i.e., no level dominance was observed.

Experiment 2 demonstrates that level dominance is reduced when the target and nontarget sounds have less shared frequencies. For experiment 2A, the noise nontarget, which was broadband, provided a release from level dominance. For experiment 2B, in the *off-freq* condition, where the target and nontarget sounds occupied wholly different frequency regions, thresholds for low- and high-level targets were approximately the same. Comparing the results for low and lower conditions across the two experiments, a release from level dominance occurred when the target and nontarget spectral extents differed, as in experiment 2A, but not when they essentially shared the spectral extent, as in experiment 1. This indicates that relatively small local differences in harmonic frequencies between targets and nontargets do not lead to changes in thresholds, which is consistent with past research (e.g., Zeng and Turner, 1992). Regarding the release of level dominance, we failed to uncover the systematic effects of changes in quality. The current results are consistent with mechanisms that are largely peripheral: changes in spectral range but not differences in the spectral shape and fundamental frequency led to changes in thresholds.

1.4 Summary

In the current study, the effects of sound segregation cues on multi-sound intensity discrimination were explored. In particular, we tested three acoustic cues known as compulsory segregation cues: changes in pitch, timbre and frequency extent. The data do not support that changes in stimuli which encourage sound segregation, in general, provide a release from level dominance. Changes in fundamental frequency and spectral shape did not release level dominance when the target and nontarget sounds shared the same frequency regions. Regarding changes in the degree of frequency non-overlap between the target and nontarget sounds, the data indicate a somewhat graded release from level dominance with reductions in the amount of spectral overlap. Moreover, makeup of the lower-level sounds did not impact intensity discrimination thresholds for the higher-level sounds, indicating the strong salience of the higher-level sounds.

Chapter 2

Covarying sensitivity between level dominance and stream segregation

2.1 Introduction

Auditory level dominance (Berg, 1990) is the tendency for more intense sounds to be overweighted, leading to suboptimal performance when integrating information across a series of sounds with alternating intensities. The first chapter of this dissertation demonstrated that a frequency non-overlap, a robust acoustic cue for stream segregation, also led to a notable reduction in level dominance measured by intensity discrimination thresholds.

Studies have examined the possibility that a shared underlying mechanism could influence performance in multiple behavioral tasks. For instance, a study in visual perception found correlated individual differences between metacognition of visual perception and visual short-term memory, suggesting a shared underlying mechanism (Samaha and Postle, 2017). Similarly, differences in performance across two auditory spatial selective tasks in normal-hearing listeners were attributed to differences in subcortical coding and attentional control (Dai and Shinn-Cunningham, 2016).

Oberfeld et al. (2012) manipulated the perceived lateralization of the masker via variation of the interaural time difference (ITD) to test whether this cue influences the intensity resolution. Their results demonstrated that a forward masker perceived as lateralized to the other side of the subsequent target reduces masking compared to a masker lateralized ipsilaterally. This supports the assumption that object segregation may promote selective attention to the target. Thus, there appears to be a connection between listeners' ability to selectively attend to low-level target sounds and their capacity to focus on one of two sound sets using segregation cues.

The direct relationship between the two abilities has been understudied. Previous work has focused on assessing the impact of various manipulations on the sensitivity to intensity changes (e.g., Schlauch et al., 1997) or the strength of stream (object) segregation (Moore and Gockel, 2002), but has not studied individual differences in both tasks jointly. As a result, a comprehensive investigation of how strongly these two measures are linked, or whether there is a shared mechanism between the two tasks, has yet to be conducted. This study is a follow-up to our earlier study (Watanabe et al., 2023). In the current study, we obtained measurements from both the level dominance task and the stream segregation task to test if variances in these two measures are related. We hypothesized that changes in sound quality with a higher tendency to cause stream segregation would also lead to less level dominance (lower intensity discrimination thresholds for low-level targets).

The current study includes two tasks: a multi-sound intensity discrimination task described in the previous chapter, and a temporal-gap duration discrimination task adopted from

Micheyl and Oxenham (2010). They described an objective measure of stream segregation showing that a gap duration discrimination threshold decreases as ΔF increases. This led us to think that this measure can be used as a proxy for the strength of stream segregation across various changes in quality. In this paradigm, two sets of sounds (target and nontarget sounds) were presented and listeners were instructed to focus on the relative timing of target sounds while ignoring the nontarget sounds (Figure 2.1). In both tasks, the makeup of the nontarget sounds was varied to influence the difference in "quality" between the target and nontarget sounds. The same sets of target and nontarget sounds were used in both tasks to investigate how a change in one type of quality change affects release from level dominance and the strength of perceptual segregation. We evaluated the strength of the correlation between thresholds measured in two tasks for each subject. The individual correlations were then analyzed collectively to evaluate the overall strength of the correlation at the group level.

2.2 Methods

2.2.1 Participants

All experimental procedures were approved by the Institutional Review Board of the University of California, Irvine. Eight subjects participated in the experiment (aged 18–60 years old, three males). The subjects were paid an hourly wage, except for K.W., the author. All subjects had absolute thresholds of 20 dB hearing level (HL) or better for audiometric frequencies between 250 and 4000 Hz. We did not exclude participants with absolute thresholds of 25 dB HL or higher for 6000 and 8000 Hz because the highest frequency used to produce stimuli in experiments was 4000 Hz.

2.2.2 Task and stimuli

The stimuli were digitally generated using a sampling frequency of 44100 Hz. The experiment was conducted in a double-walled, sound-attenuated booth. For the level dominance task, the paradigm was similar to the low target (LT) stimulus of Watanabe et al. (2023). The sole difference was that the initial value of ΔL was 4.5 dB.

The other task, the gap duration discrimination task, was adapted from Micheyl and Oxenham (2010). Figure 2.1 shows the schematic of a trial. The stimuli consisted of sequences of target and nontarget sounds. Both target and nontarget sounds were presented at 65 dB SPL. In this experiment, listeners' task was to focus on the target sound stream and indicate whether the last target sound was "early" or "late". Each stimulus presentation was followed by visual feedback as to the correctness of the subject's response. In each trial, a target stream was 8 repetitions of 50-ms sound pips with 10-ms cosine squared onset and offset ramps, separated by 300 ms until the last inter-sound interval. The last target sound was shifted earlier or later (equal odds) by Δt, where Δt was varied adaptively using a two-down one-up procedure (Levitt, 1971). The initial value of Δt was 50 ms, with the initial step size of 10 ms. After two reversals, the step size was reduced to 5 ms, and after the following two reversals, the step size was reduced to 2.5 ms and the track continued until an additional six reversals were completed. The 71 % correct threshold is the arithmetic mean of the last six reversals. Another set of sounds, nontarget sounds (50 ms long, 10-ms cosine-squared onset and offset ramps), were inserted between the target sounds to distract the listeners, making the detection of a temporal shift in the last target sound more challenging. The onset of the

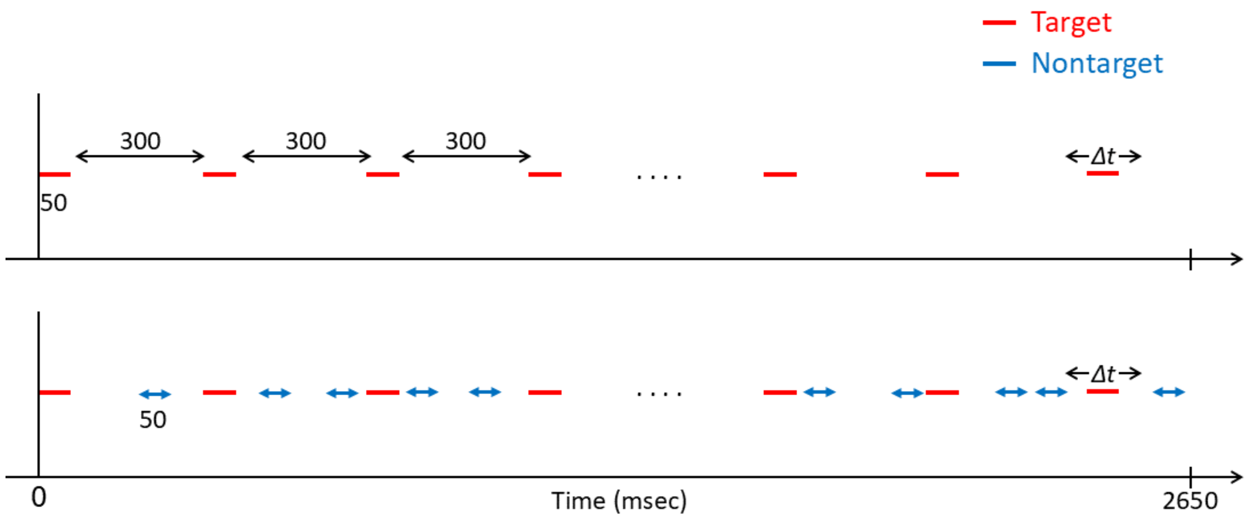

Figure 2.1: Schematic of example stimulus sequences presented on a trial of stream segregation experiment (adopted from Micheyl and Oxenham, 2010). Listeners initially practiced and ran the *baseline* condition (Top), where only the target sounds were presented. The duration of the interval between consecutive target sounds was constant (300 ms) except for the last pair, where it was reduced or increased (with equal probability) by Δt. Then nontarget sounds were inserted between the target sounds (Bottom). The onsets of the first and second nontarget sounds were randomly selected from uniform distributions over the ranges of 0 to 100ms and 150 to 250ms, respectively.

first nontarget sound was randomly selected from a uniform distribution over the range of 0 to 100 ms, ensuring that it occurred within the first half of the interval. The onset of the second nontarget sound was randomly selected from a uniform distribution over the range of 150 ms to 250 ms, placing it within the second half of the interval. This design ensured that the nontarget sounds were distributed within the target interval without overlapping with the target sounds. One nontarget sound was presented between the first and second target sounds, two nontargets between each pair of subsequent targets, and another nontarget after the last target. The logic behind this paradigm is that the more perceptually dissimilar the target and nontarget sounds are, the more likely the target sounds will form their own stream, thus allowing listeners to better detect the temporal shift of the last target sound. In the *baseline* condition, only the target sounds were present.

Four listeners ran the level dominance task and then gap duration discrimination task; the

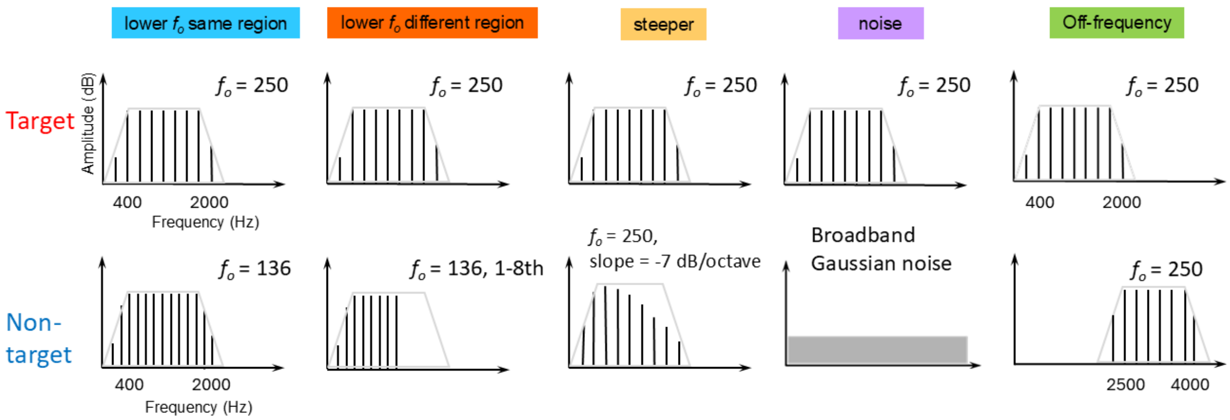

Figure 2.2: Description of sound stimuli and conditions in the experiment. The target sounds were the harmonic complexes of f_o = 250 Hz. The characteristics of the nontarget sounds varied to form five conditions: change in f_o associated with a change in frequency region, *noise*, and *off-frequency*.

order was the opposite for the other four listeners.

Stimuli used in the current study were similar to those from Watanabe et al. (2023), as shown in Figure 2.2. The bandpass filters used to produce stimuli in this experiment were 6th-order Butterworth filters. Note that for the rest of the section, stimuli were bandpass filtered with cutoff frequencies of 400 and 2000 Hz, unless stated otherwise. The target sound was always a harmonic complex of 250 Hz fundamental frequency (f_o) with a flat spectrum (spectral slope = 0 dB/octave). In the *lower* f_o*, same region* condition, nontargets had the f_o = 136. In the *lower* f_o*, different region* condition, the nontarget sound was the 1st - 8th harmonics of f_o = 136. This set of harmonic numbers was chosen to roughly match the harmonic strength of the target sound (2000/250 = 8) while introducing a non-overlapping frequency region. In the *steeper* condition, nontarget was a bandpass filtered harmonic complex of f_o = 250, and the slope of the spectrum fell at a rate of 7 dB/octave. The *noise* condition was the bandpass filtered Gaussian noise; finally, in the *off-freq* condition, the nontargets were the harmonics of a 250-Hz fundamental and bandpass filtered with cutoff frequencies of 2500 and 4000 Hz, which equals to 10th – 16th harmonics.

The makeup of the target and nontarget sounds was identical across the level dominance and stream segregation tasks.

2.2.3 Procedure

The protocol for the level dominance task was identical to that in the previous chapter. The protocol for the stream segregation task was as follows. Listeners initially practiced the *baseline* condition to ensure they understood the task. Next, six threshold estimates were collected for the *baseline* condition. After completion, the listeners ran the remaining conditions in random orders (counterbalanced). The order of running these conditions was never identical to that of the level dominance task. For the initial non-*baseline* condition encountered by a listener, nine threshold estimates were collected; otherwise, six threshold estimates were collected. For one listener, only five threshold measurements per condition were obtained due to their time constraint.

2.2.4 Correlational analysis

We computed each subject's Pearson correlation coefficient (r) using the data points from five non-*baseline* conditions. Although the Pearson correlation is traditionally not ideal for assessing the strength of correlation in within-subject data due to the potential influence of covariance from repeated measures, we justified its use in the current study. By employing a counterbalanced design, we minimized the impact of covariance associated with the repeated measures.

To assess the influence of individual correlations, we employed Fisher's z-transformation method (Silver and Dunlap, 1987; Corey et al., 1998) for combining and averaging Pearson coefficients. First, each subject's Pearson correlation coefficient (r_i) was transformed into a Fisher's z-score (z_i). Fisher's z-transformation is defined as:

$$z_i = artanh(r_i) \tag{2.1}$$

After z-transformation, we calculated the mean Fisher's z-scores across subjects. This mean z-score was then converted back to an average Pearson correlation coefficient ($\bar{r}$) using the inverse Fisher transformation:

$$\bar{r} = tanh(\bar{z}) \tag{2.2}$$

To assess whether the average correlation significantly positive, we conducted a one-tailed, one-sample t-test on the mean Fisher's z-score.

2.3 Results

2.3.1 Level dominance task (multi-sound intensity discrimination)

Figure 2.3 plots the mean thresholds for all conditions: *baseline*, *steeper*, *lower f_o same region*, *lower f_o different region*, *noise* and *off-freq*. Non-*baseline* conditions are plotted in the expected decreasing order of magnitude based on Watanabe et al. (2023). A one-way repeated measures ANOVA was performed to investigate the effect of the five non-*baseline* conditions on threshold measurements. Mauchly's Test indicated a violation of the sphericity assumption, $\chi^2(9) = 0.04$; therefore, the Greenhouse-Geiser correction was applied (ϵ

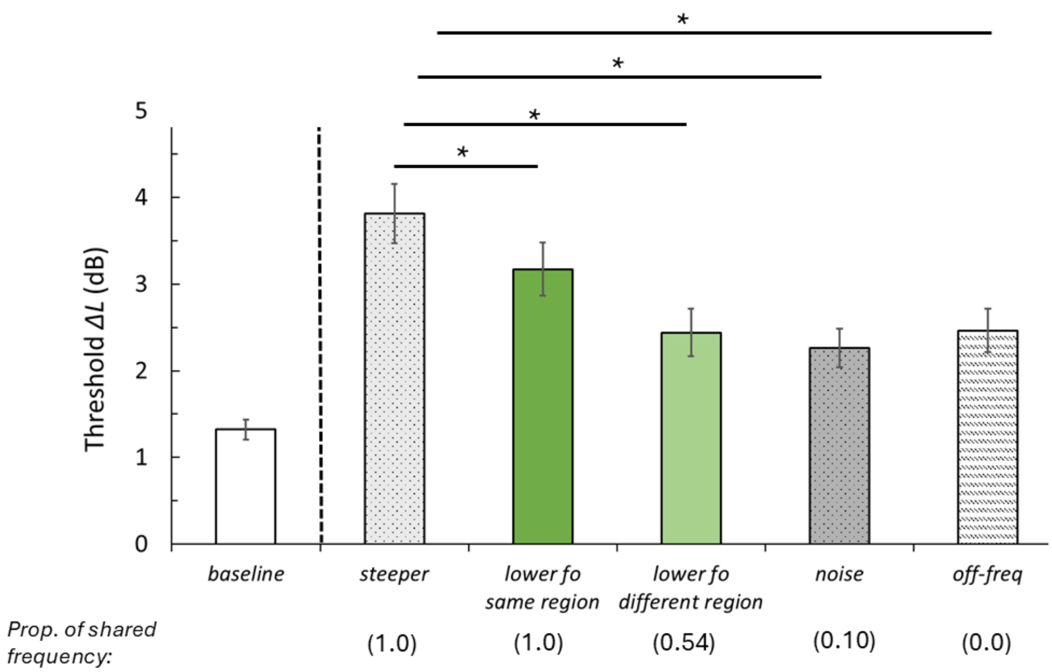

Figure 2.3: Thresholds for level dominance task as ΔL in dB. Error bars indicate $\pm$ one standard error of the mean. Statistical analyses are described in the text. The values at the bottom indicate the proportion of shared frequency. See texts for details.

= 0.44). The ANOVA revealed a significant main effect, $F(1.8,\ 12.3) = 21.6$, $p < .01$. We conducted planned comparisons between the *steeper* condition with other non-*baseline* conditions. One-tailed paired t-test after applying a Bonferroni correction for multiple comparisons (adjusted alpha = 0.0125) showed significant differences between *steeper* and *lower f_o same region*, *lower f_o different region*, *noise*, and *off-freq* conditions. ($p < .01$). Similarly as in Chapter 1 experiment 2, we quantified the proportion of shared frequencies between target and nontarget sounds as follows: *steeper* (1.0), *lower f_o same region* (1.0), *lower f_o different region* (0.54), *noise* (0.10), and *off-freq* (0.0).

Consistent with Chapter 1, the threshold associated with a change in the spectral slope of the nontarget sounds demonstrates a large degree of level dominance. The next largest threshold, but significant release from *steeper* condition, is observed when the nontargets involved a change in pitch but occupied the same frequency region as the targets (*lower f_o*

same region). An even further reduction in thresholds was observed in the remaining three conditions involving changes in frequency overlap. When a change in pitch is associated with a change in frequency extent, that leads to a reduction in threshold, as well as *noise* nontargets.

Lastly, thresholds for *off-freq* condition show some release; however, not only does some level dominance remain, but it also fails to provide the largest release compared to all other conditions. The potential reason for this discrepancy from the earlier study is the smaller degree of frequency non-overlap between targets and nontargets in the current study. In the earlier study, targets had the highest frequency of 1750 Hz, while the nontargets had the lowest frequency of 2500 Hz, resulting in a frequency non-overlap of 750 Hz. However, in the current study, the difference between the upper cutoff frequency of targets and the lower cutoff frequency of nontargets was 500 Hz. Another potential reason for the inconsistency is that the subjects were undertrained and lacked sufficient practice before proceeding to the main data collection.

2.3.2 Gap duration discrimination/Stream segregation task

Figure 2.4 plots the mean gap duration discrimination thresholds. A lower threshold is associated with a greater tendency for parallel sounds to form separate streams (Micheyl and Oxenham, 2010). The Δt thresholds in the *baseline* condition are about 15 ms. Adding nontarget sounds with a steeper spectral slope caused the largest elevation in thresholds among the five tested conditions, parallel to thresholds in the level dominance task above. This time, a notable reduction in threshold was observed only when the change in pitch was associated with a change in frequency overlap. The thresholds for *noise* and *off-freq*

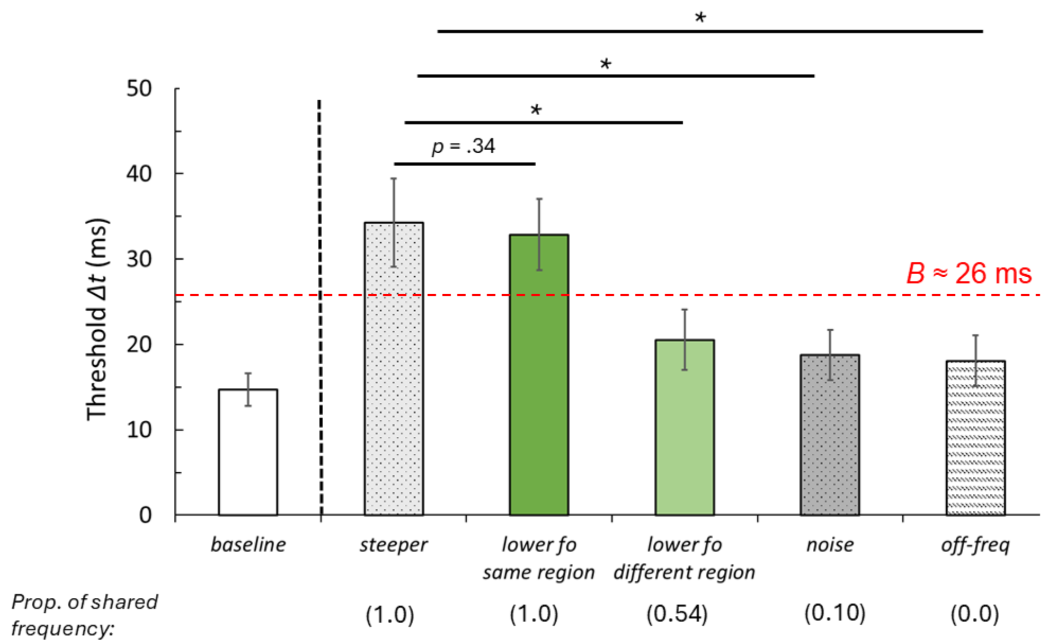

Figure 2.4: Thresholds for gap duration discrimination task as Δt in ms. The same set of conditions as in the level dominance task was tested. Error bars indicate $\pm$ one standard error of the mean. Statistical analyses are described in the text. A red dotted line indicates the idealized local boundary B computed from a computer simulation (See Appendix A).

conditions are comparable to *lower* f_o *different region*.

One-way ANOVA revealed a significant main effect, $F(4{,}28) = 16.49$, $p < .01$. One-tailed paired t-test after applying a Bonferroni correction for multiple comparisons (adjusted alpha = 0.0125) did not show a significant difference between *steeper* and *lower* f_o *same region*, ($p = .34$) but the difference was significant between *steeper* and *lower* f_o *different region*, *noise*, and *off-freq* conditions. ($p < .01$). The red dotted line indicates B computed from a computer simulation (See Appendix A). Given the amount of randomization, the threshold a noiseless observer can achieve by only focusing on the last triplet of the sound sequence is: $B \approx 26$ ms. This indicates that for $\Delta t \geq 26$ ms, local comparison cue may be contributing to the performance, but the conditions with threshold $\Delta t < 26$ ms (*lower* f_o *different region, noise and off-freq*) indicate that local comparison is not the sole cue and some additional cue (presumably within-target comparison) is incorporated to achieve those thresholds.

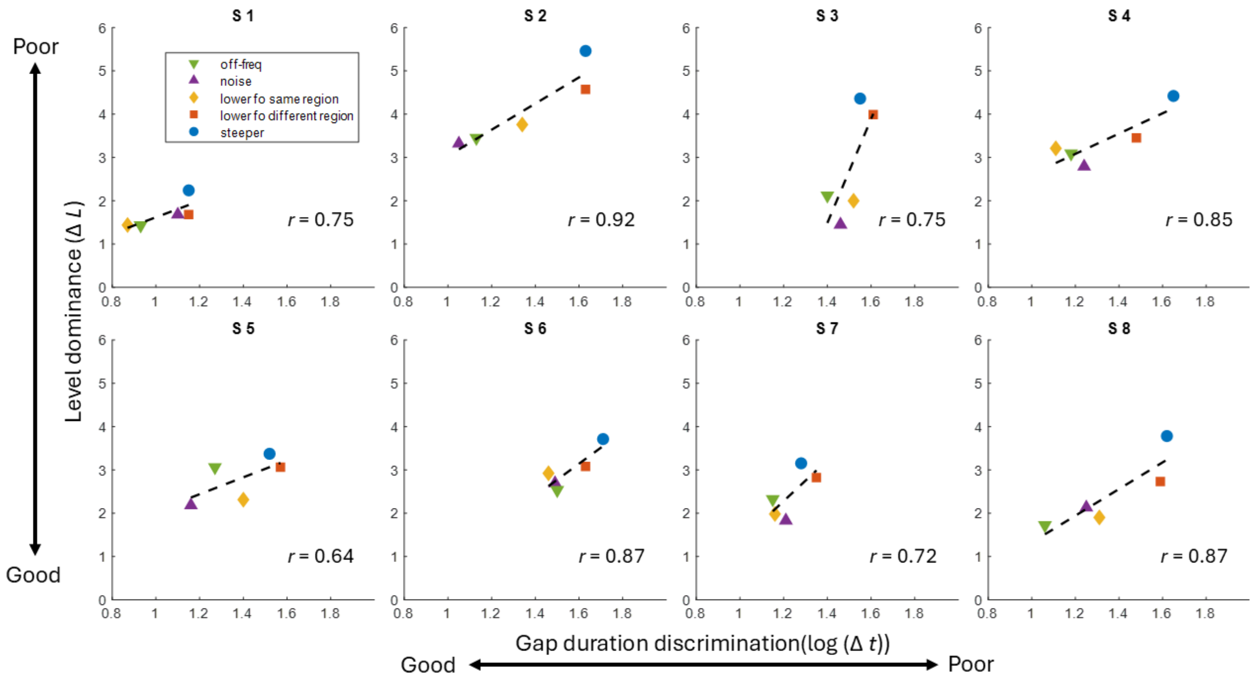

Figure 2.5: Individual scatterplots for eight subjects showing the relationship between threshold measurements from the gap duration discrimination task (log Δt in ms; x-axis) and the level dominance task (ΔL in dB; y-axis). Data points are color-coded as indicated in the legend to represent different conditions. Each scatterplot includes a best-fit linear line to illustrate the trend for each subject and displays the Pearson correlation coefficient (r) computed for each individual.

2.3.3 The relation between level dominance and stream segregation tasks from individual data

Scatter plots in Figure 2.5 describe the joint relation for thresholds in level dominance and gap duration discrimination tasks. Thresholds for the latter are plotted as log Δt (e.g., Abel, 1972). Symbols are color-coded to indicate different conditions. It is notable that two conditions with higher mean thresholds in both tasks (*steeper* and *lower f_o same region*) are located at the top right of each scatter plot. For each subject, a linear regression was performed. The estimated Pearson's product correlation is also indicated for each subject.

Individual correlation coefficients range from 0.64 to 0.92. The average correlation (averaged via following a Fisher's z-transformation) is: $\bar{r} = 0.81$, indicating a strong positive correlation. The one-tailed one-sample t-test shows that this average correlation is significantly positive (greater than 0), $t(7) = 4.551$, $p < .01$. The 95% one-tailed confidence interval is: $\bar{r} \geq 0.581$.

The results suggest the following: there is a general trend among individuals that if the sound manipulation is a strong cue for stream segregation, that also tends to be a strong cue for reduction in level dominance. This implies that the two tasks may share common underlying mechanisms or factors that influence performance.

2.4 Summary

This follow-up study from Chapter 1 evaluated the relation between a level dominance task and a gap duration discrimination task (a proxy for stream segregation, Micheyl and Oxenham, 2010) for five of the stimuli tested in Chapter 1. The data demonstrate that introducing frequency non-overlap provides a strong cue that facilitates performance for both tasks.

Averaging the individual correlation coefficients using Fisher's z transformation to examine the relationship between the two tasks demonstrated that the performance in one task is highly correlated with the performance in the other. Specifically, sound manipulations facilitated (or did not facilitate) performance in the level dominance task similarly to how they did in the stream segregation task. For instance, using Gaussian noise nontargets against harmonic complex targets resulted in both a significant reduction in level dominance and a strong tendency for target and nontarget sounds to be perceptually segregated. In contrast, a timbral cue (a change in spectral slope), did not effectively reduce level dominance or promote segregation. We observe a hierarchy among compulsory segregation cues (Moore and Gockel, 2012).

By correlating the measurements representing the degree of stream segregation and the release in level dominance, we were able to observe not only the overall relation between the two but also within-subject effects. We could determine that the degree of stream segregation is systematically related to the release from level dominance observed in the same subjects.

The results of this chapter suggest a common mechanism underlying both the level dominance and stream segregation tasks. However, studies involving computational modeling may be necessary to further test this hypothesis.

Chapter 3

Attentional direction and level dominance: insights from an EEG study

3.1 Introduction

Humans are constantly exposed to multiple sources of information, facing the need to extract the most relevant signals while filtering out irrelevant information. Auditory selective attention plays a crucial role in this process. When sounds exhibit acoustic regularities, the alignment of neural oscillations to stimulus temporal structures, (i.e., neural entrainment; see Ding and Simon, 2012), may facilitate this process.

Laffere et al. (2020) studied the individual difference in "direction of attention", defined as the ability of listeners to selectively attend to specific time intervals at which a target

Figure 3.1: Schematic representation of the stimulus used by Laffere et al. (2020). Tones that belong to the same band are indicated in orange.

appears in a sound sequence. They used long-repeating stimuli consisting of two “bands” completely non-overlapping in time and frequency (Figure 3.1). Listeners were instructed to listen to one band at a time and perform a task requiring selective attention. Both bands contain regularities at the same rate but are separated in time by half a phase cycle (180°) at the within-band presentation rate. They discovered that individual differences in temporally selective attention manifest in the neural phase angle derived from EEG data. Successful selective attention to both bands led to a bigger phase shift between the conditions. An MEG study by Srinivasan and Petrovic (2006) reported that presenting two types of flickers 180° out of phase causes sensors located over frontal and some occipital areas to exhibit 180° phase shift between conditions.

On the other hand, certain stimuli capture attention involuntarily, disrupting voluntary attention (Remington et al., 1992). Schröger (1996) demonstrated that deviant auditory stimuli can automatically capture attention and elicit the mismatch negativity (MMN), resulting in reduced capacity for processing subsequent stimuli. Moreover, the strength of attentional capture can vary with the physical salience of the stimuli. In a visual search task, salient distractors, such as those with abrupt onsets or high brightness, demonstrate stronger early perceptual processing and require more attentional resources to be suppressed later (Chen et al., 2023). Likewise, changing sound intensity may be an effective way to manipulate involuntary attention. Previous studies hint that the presence of louder elements causes failure in selective attention to softer elements (Oberfeld et al., 2013) or attention is directed to the louder elements (Turner and Berg, 2007).

The current study aims to explore how relative levels of sequential sounds may influence attention to predictable temporal patterns. Our main interest lies in how task-irrelevant but salient stimuli interfere with the process of selective attention. We hypothesize that the degree to which listeners can attend to target-time intervals is linked to neural phase angle, as shown by previous studies (Laffere et al., 2020; Laffere et al., 2021).

3.2 Methods

3.2.1 Participants

All experimental procedures were approved by the Institutional Review Board of the University of California, Irvine. A total of 16 participants (age range: 18 to 23) were recruited for the study. All but one participant had absolute thresholds of 20 dB HL or better for audiometric frequencies between 250 and 8000 Hz. One participant had an absolute threshold of 25 dB HL at 6000 Hz on their right ear. Listeners were recruited via posters placed around the UC Irvine campus.

3.2.2 Stimulus design

The sounds were digitally generated with a sampling frequency of 44100 Hz using MATLAB (Mathworks, Inc) and the Psychophysics toolbox (Brainard and Vision, 1997). The stimuli were presented to participants via Focusrite Scarlett 2i4 USB audio interface (High Wycombe, United Kindgom) and an electrostatic Stax SRS-002 In-The-Earspeaker System (Stax Ltd., Japan).

Figure 3.2 is a schematic of the stimulus design, plotting sound level as a function of time. The stimuli were 75-ms cosine-ramped harmonic complexes bandpass filtered at 400-2000 Hz (fundamental frequency described later) followed by an inter-stimulus interval (ISI) of 175 ms. Therefore, inter-onset intervals (IOI) between each stimulus were 250 ms, with a *stimulus presentation rate* of 4 Hz. Stimuli consisted of complex tones alternating between two different base fundamental frequencies ($f_o Base$): 250 Hz and 180 Hz. Specifically, the odd-numbered components had a base fundamental of 250 Hz ($f_o Odd$), while the even-numbered components had a base fundamental of 180 Hz ($f_o Even$). It has been demonstrated by Watanabe et al. (2023) that two sounds with different f_o's sharing the spectral region (both filtered at 400-2000 Hz) do not cause release from level dominance.

We presented a sound sequence consisting of three sets of odd and even components, resulting in a total of six components per trial. For half of the session, listeners were instructed to attend to odd components ("Attend Odd" condition) and attend to even components ("Attend Even" condition) for the other half of the session. Sounds that listeners are instructed to listen to are referred to as "targets", and ones they are instructed to ignore are referred to as "nontargets".

Within each trial, the fundamental f_o for each component was chosen randomly from one of the following: the base fundamental frequency, $f_o Base$, or a slightly incremented fundamental, $f_o Base + \Delta f_o$, where $\Delta f_o / f_o Base = 5\%$. In other words, the chosen f_o for odd components could be either 250 Hz or 250 + (0.05*250) = 262.5 Hz. Similarly, for even components the fundamental could be either 180 Hz or 189 Hz. For ease of description,

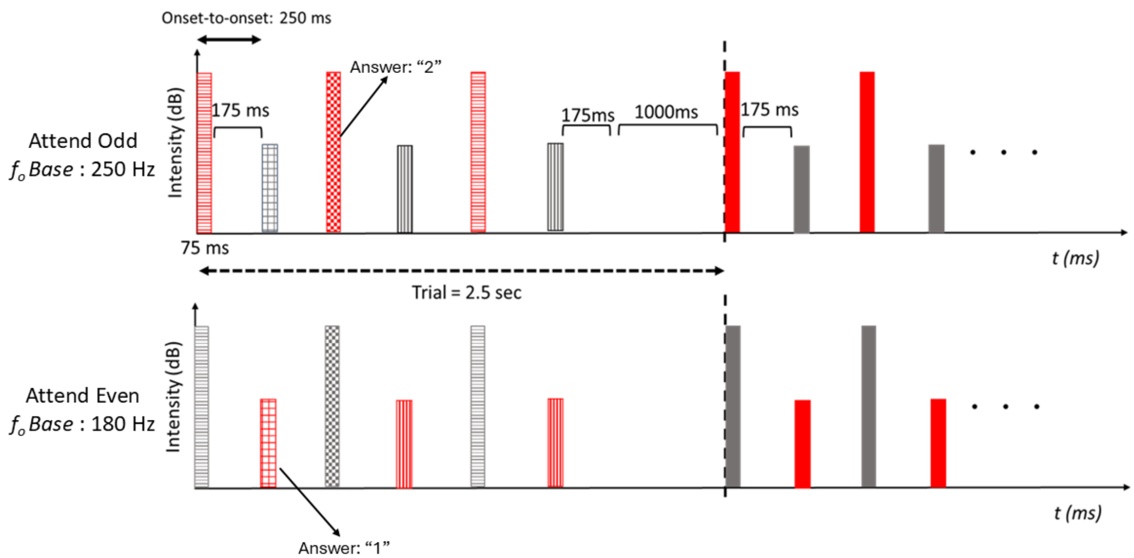

Figure 3.2: Schematic representation of the stimulus design. Bars in red indicate the sounds to which listeners are asked to attend ("targets), and ones in gray are the sounds listeners are to ignore ("nontargets"). Listeners are instructed to respond which of the three targets has a different f_o from the other two. In the example above, the correct response is "2" in Attend Odd but "1" in Attend Even, illustrated by a different pattern.

the higher f_o of the two is referred to as "high pitch" and the lower one is referred to as "low pitch". This $\Delta f_o/f_o$ value was chosen based on past work on fundamental frequency discrimination (Ruggles et al., 2014; Madsen et al., 2017) to ensure that most listeners would not have difficulty discriminating between the two possible fundamental frequencies. However, for three listeners who reported that 5% was not easily discriminable, the difference was increased to 7%. The task is described in the next subsection.

The IOI from one sound to another within the target (i.e., odd to odd, even to even) is 500 ms, so the *within-target presentation rate* is 2 Hz. Accordingly, any adjacent target sounds were separated in time by half a phase cycle (180°) at 2 Hz. A single 2.5-s trial comprises three presentations of odd and even components, followed by a 1-second silence. Trials were concatenated together into blocks of 35 trials, and there was one run of 3 blocks, a total of 105 trials per condition.

To investigate the effects of different sound levels, the following three level pairs were tested. 80 vs 30, 55 vs 55, and 30 vs 80. For example, 80 vs 30 indicates levels of 80 dB SPL for odd components and 30 dB SPL for even components. Figure 3.2 is an example illustrating the 80 vs 30 condition. As mentioned above, Attend Odd and Attend Even conditions were conducted for each level pair, resulting in a total of 6 conditions. For example, *Attend Odd: 80 dB* indicates that listeners are asked to attend to target odd components presented at 80 dB SPL, while nontarget even components are presented at 30 dB SPL.

3.2.3 Procedure

Participants performed a three-interval forced choice (3IFC) fundamental frequency discrimination task (Figure 3.3). In the Attend Odd and Attend Even conditions, participants were instructed to selectively listen to target sounds with corresponding temporal positions and respond which of the three targets has a different pitch from the other two. For example, if the pitch pattern of targets in one trial is "high-low-high", then the correct response is to press a "2" key. If three targets are assigned the same pitch ("high-high-high" or "low-low-low"), the correct response is to wait until the next trial begins without pressing any key. The stimulus was designed in a way that attending both targets and nontarget sounds has no advantage over attending just targets. The latency window for a response to be recorded was from 175 ms after the offset of the last sound in the sequence to 150 ms before the beginning of the next trial. Feedback is presented on a computer display to notify subjects of correct, and incorrect responses. Behavioral performance is measured as percent correct. Listeners first practiced in the 'only targets' condition, where nontargets were absent. We tested both 80 dB and 30 dB SPL levels in this setup and confirmed that all listeners achieved above 80% accuracy at both levels. This indicates that the sound level of the targets itself does

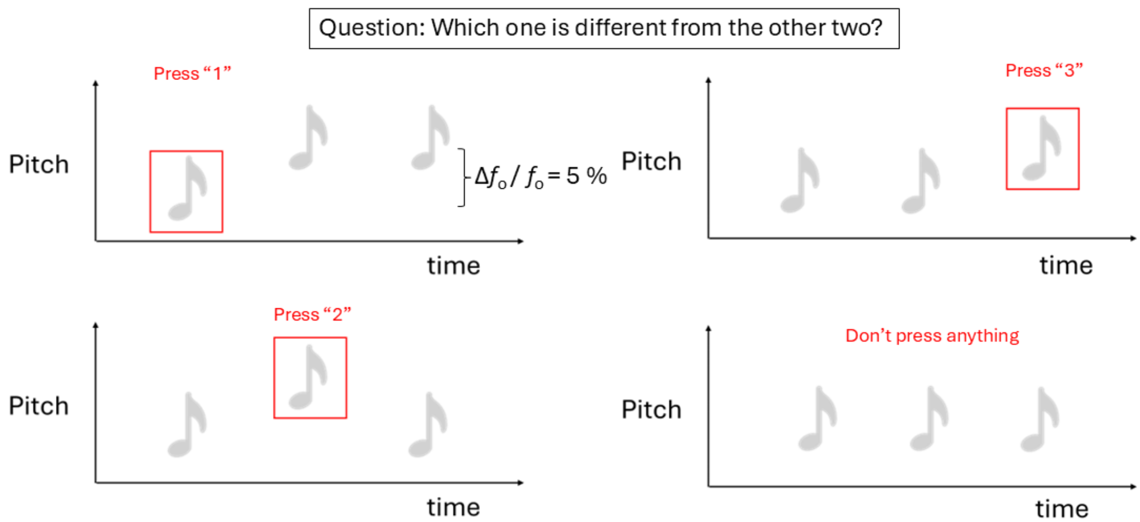

Figure 3.3: Schematic representation of the task. Listeners were instructed to selectively listen to target sounds and detect which of the three targets has a different pitch from the other two. If three targets had the same pitch, the correct answer is not to press any key. The proportion $\Delta f_o/f_oBase$ was set to 5%, but it was raised to 7% for three listeners.

not influence performance. Then they proceeded to the main data collection. The condition order was counterbalanced across listeners. Eight subjects completed Attend Odd and then Attend Even, and the order was the opposite for the other eight listeners. The order of the level pair was also randomized.

3.2.4 EEG data processing and analysis

High-density EEG (128 channels) was recorded at a sampling rate of 1000 Hz and filtered offline with a passband of .7 – 30 Hz. All channel impedances were kept below 20 kΩ. EEG data were segmented into 1.5-sec epochs aligned with the onset of the trial (without the 1-second silence). Sources of artifacts such as eye blinks and eye movements were identified by independent component analysis (ICA; Hyvärinen and Oja, 1997) of the epoched recording. EEG data were processed to extract two neural measurements of interest: average neural phase and inter-trial phase coherence (ITPC). To compute average neural phase and ITPC, a Hanning-windowed fast Fourier transform is conducted on each epoch to extract the phase for each trial at 2 Hz. The amplitude of the resulting complex vectors was then set to one by dividing by the vector's length. The vectors were averaged, and the length of the resulting vector was calculated as ITPC, while the phase of the resulting vector equaled the average phase across trials.

Average phase angle indicates the timing of the neural response with respect to the stimulus. ITPC is a measure of the consistency of alignment of neural phase, serving as a reliability indicator for the average phase angle. This measure varies from 0 (no consistency) to 1 (perfect alignment). Following Delorme and Makeig (2004), ITPC was calculated as:

$$ITPC(f,t) = \frac{1}{n}\sum\nolimits_{k=1}^{n} \frac{F_k(f,t)}{|F_k(f,t)|} \tag{3.1}$$

where t is time, f is frequency, F is the Fourier transform, and n is the number of trials. Average phase angles and ITPC were computed for Attend Odd and Attend Even in each level pair. Descriptive and inferential analyses of neural phase data were carried out using the MATLAB built-in functions and circular statistics toolbox (Berens, 2009).

To select the appropriate channels for analyses, we employed a multistep process to ensure that channels best represented the neural activity of interest. First, any channel with an ITPC value less than 0.1 (Laffere et al., 2020) at 2 Hz for the conditions; *Attend Odd: 80 dB* and *Attend Even: 80 dB* was excluded from further analyses. This exclusion criterion was applied because, in these conditions, the temporal position of the loud aligns with the target time intervals. Channels with poorly aligned phases (low ITPC) under these conditions are thus considered unreliable measures. Second, reliable channels were identified by their ITPC at 2 and 4 Hz, frequencies relevant to the stimulus structure. Specifically, any channel with

ITPC at these relevant frequencies (2 and 4 Hz) lower than the ITPC at adjacent frequencies (1.33 Hz, 3.33 Hz, and 5.33 Hz) was excluded from further analysis. Finally, we computed the distance in average neural phase angles between the *Attend Odd: 80 dB* and *Attend Even: 80 dB* conditions using the 'circ_dist' function Berens (2009). The channel exhibiting the greatest distance in average neural phase angles was selected for further analysis. This inclusion criterion was applied because the temporal positions of loud components in these two conditions are shifted by 180°.

3.3 Results

3.3.1 Behavioral data

Figure 3.4 (Top) plots the percent correct responses of a fundamental frequency discrimination task as a function of level pairs and attending conditions. For ease of description, the conditions *Attend Odd: 80 dB* and *Attend Even: 80 dB* are together referred to as 'Attend Loud' conditions. Similarly, the conditions *Attend Odd: 55 dB* and *Attend Even: 55 dB* are referred to as 'Attend Equal' conditions, while the other two conditions are referred to as 'Attend Soft.' The percent correct responses in three pairs of conditions are averaged and plotted as Figure 3.4 (Bottom).

The chance level was 25 % as there were four possible choices in every trial. As expected, listeners achieved the highest accuracy in two conditions where the instruction was to attend to louder targets in the presence of softer nontargets. However, it was somewhat surprising that the performance for Attend Equal was comparable to or even slightly lower than Attend Soft (six out of sixteen demonstrated such a pattern). This is inconsistent with the earlier

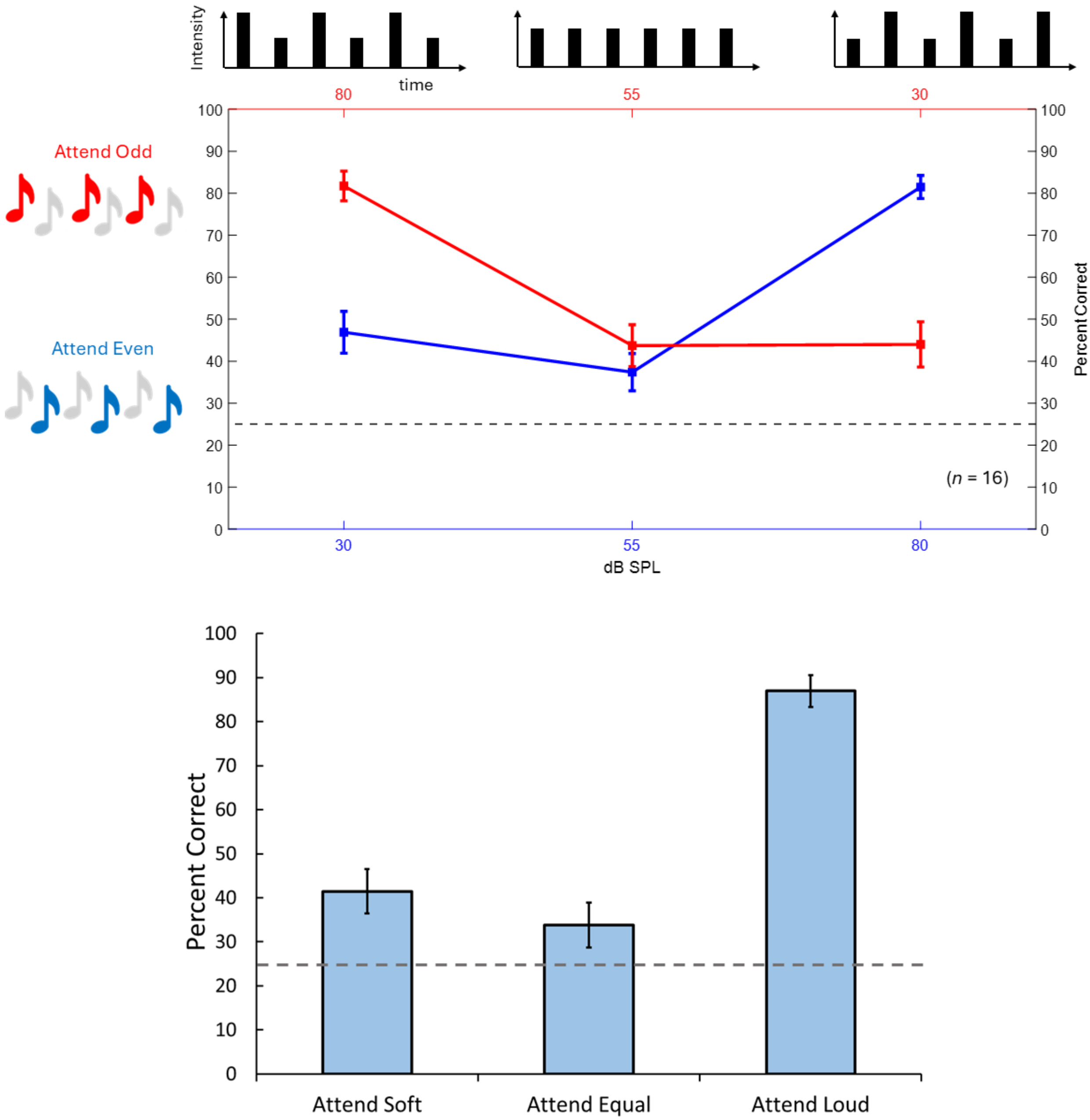

Figure 3.4: Behavioral data. The top plot shows the percent correct as a function of level pairs and attending conditions. The bottom plot shows the average percent correct for Attend Loud, Attend Equal, and Attend Soft. Error bars indicate ± one standard error of the mean.

finding that listeners can achieve decent performance in a similar task where sounds are presented at equal levels (Laffere et al., 2020; Laffere et al., 2021). The potential reason for inconsistency will be discussed later. Because reliable selective attention performance was not observed in Attend Equal conditions, we will only focus on Attend Loud and Attend Soft conditions for the rest of the chapter.

3.3.2 EEG data: analyzing neural phase angles

The polar plots (Figure 3.5 and 3.6) show the average neural phase angles, with the radii representing ITPC. Each line corresponds to an individual subject, plotted across four conditions from Attend Loud and Attend Soft. The polar plots are organized in such a way that those lined up vertically represent the same stimulus structure (intensity pattern), while those aligned horizontally indicate the same temporal position of attention (odd or even).

Two sets of polar plots are shown. In the first set (Figure 3.5), the average phase angles for the *Attend Odd: 80 dB* condition are set to 0°, and the average phase angles for all other conditions are rotated relative to this reference. This manipulation allows for a direct comparison of phase angle shifts across different conditions, using *Attend Odd: 80 dB* as the baseline. The polar plot for the *Attend Even: 80 dB* shows that neural phase angles are clustered around 180° (average across participants = 183.9°), validating that channel selection was appropriate. The most meaningful comparison in the first set is between *Attend Odd: 80 dB* and *Attend Even: 30 dB* conditions; although they have identical intensity patterns, the temporal position of attention is shifted in time by half a phase cycle (180°) at 2 Hz.

Therefore, if listeners can direct their attention to specified temporal positions, then neural phase angles for *Attend Even: 30 dB* condition would show a 180° phase shift relative to

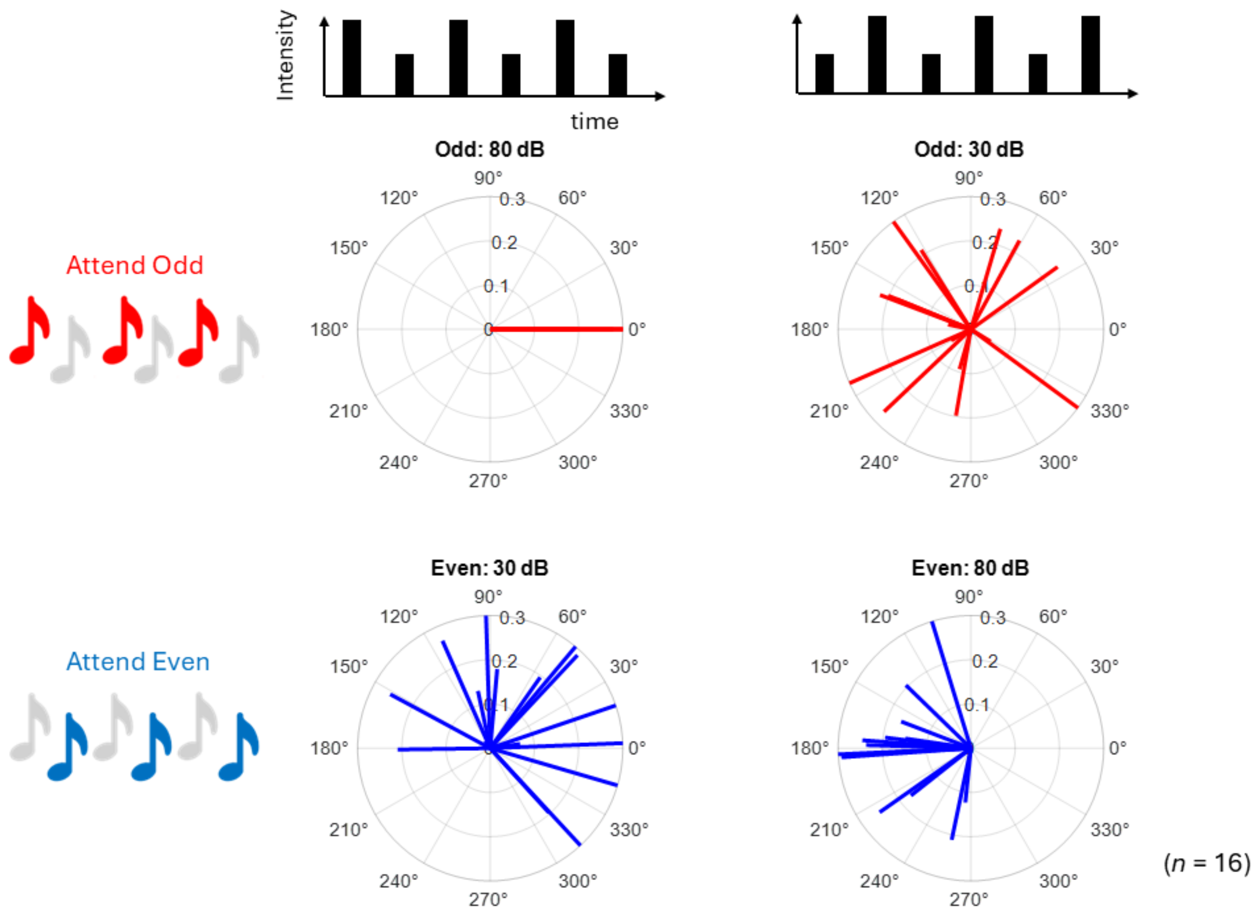

Figure 3.5: Average neural phase angle for each participant at 2 Hz relative to *Attend Odd: 80 dB*. Polar plots lined up vertically represent the conditions in which the stimulus had the same intensity pattern, while those aligned horizontally represent the conditions where listeners attended to the same temporal position.

Attend Odd: 80 dB. However, ten out of sixteen phase angles landed on the right side of the polar coordinate (average across participants = 56.4°). We performed a nonparametric test by applying the binomial test (Zar, 1999), assuming the chance of obtaining average neural phase angle: $\pi/2 < \theta < 3\pi/2$ (on the left side of the polar coordinate) is 0.5, but we did not observe a significant effect [$z = 0.375$, $p = .455$]. Therefore, we do not have sufficient evidence to reject the null hypothesis that phase angles in *Attend Even: 30 dB* are equal to 0°. Additionally, a comparison between *Attend Even: 30 dB* and *Attend Even: 80 dB* would reveal whether listeners were capable of allocating their attention to the even components consistently, regardless of the intensity arrangement. The Watson-Williams test revealed that phase patterns in two conditions are significantly different [$F(1,31) = 31.98$, $p < .001$]. It indicates that there was a significant phase shift between the two conditions, implying that the way listeners selectively attended to the even components was different between intensity arrangements.

Similarly, in the second set (Figure 3.6), the same manipulation is applied, except using the *Attend Even: 80 dB* as a baseline. The polar plot for the *Attend Odd: 80 dB* shows that neural phase angles are clustered around 180° (average across participants = 176.1°). Likewise, *Attend Odd: 30 dB* and *Attend Even: 80 dB* have identical intensity patterns but with the temporal position of attention shifted by half a phase cycle at 2 Hz. We tested if neural phase angles in *Attend Odd: 30 dB* show significant phase shifts relative to *Attend Odd: 80 dB* (significantly different from 0°) using a nonparametric test, but we did not observe a significant effect [average = 20.0°, $z = 0.25$, $p = .923$]. Also, the Watson-Williams test revealed that phase patterns between *Attend Odd: 30 dB* and *Attend Odd: 80 dB* are significantly different [$F(1,31) = 22.09$, p $< .001$], demonstrating that the way listeners attend to odd components was also inconsistent in two conditions.

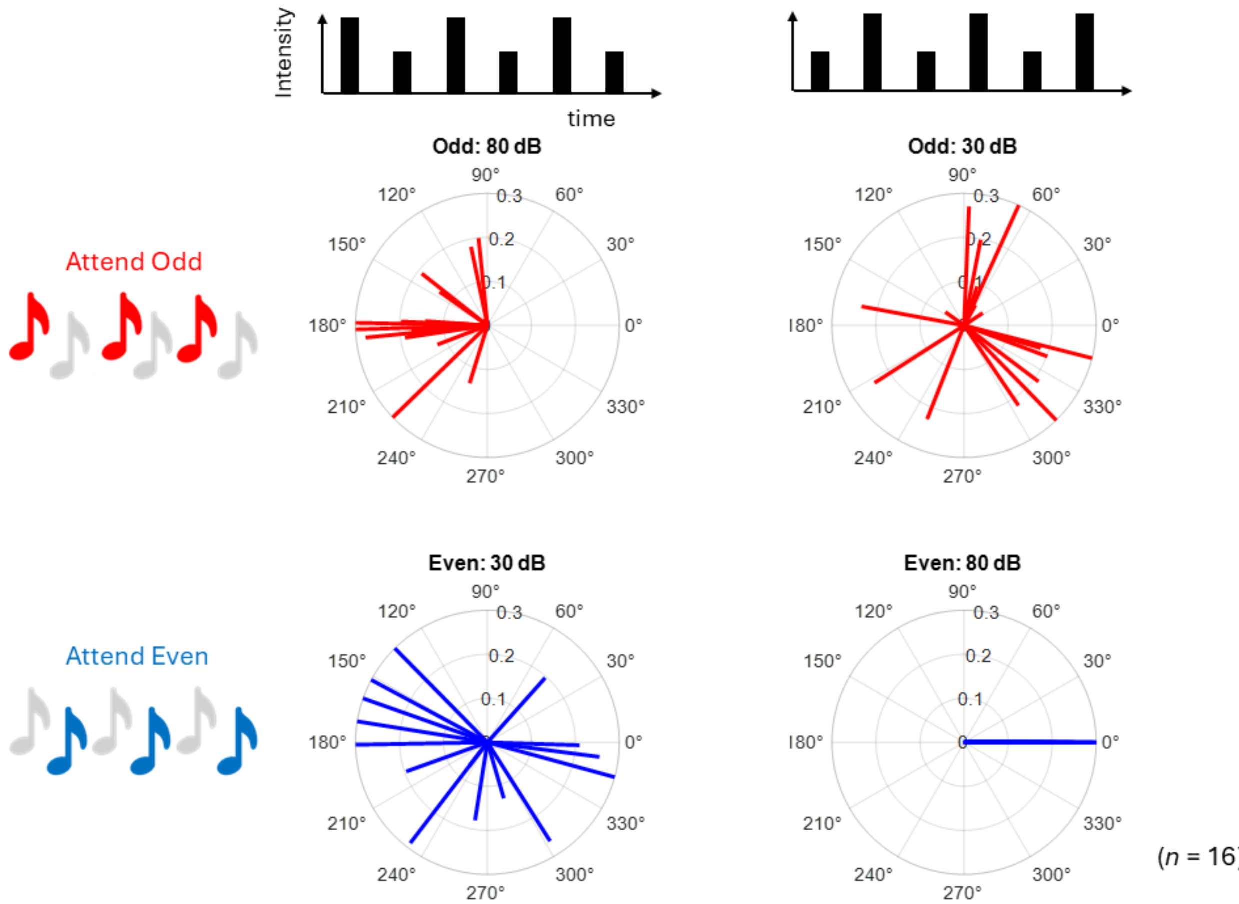

Figure 3.6: Average neural phase angle for each participant at 2 Hz relative to *Attend Even: 80 dB*.

3.4 Summary and discussion

In this study, we employed a stimulus alternating between different levels to examine the effects on auditory selective attention. We used a stimulus intensity pattern of loud-soft-loud-soft-loud-soft in one set of conditions, and listeners were instructed to attend to the louder components and perform a task in one condition, and to attend to the softer components in another condition. We repeated this procedure with a reversed stimulus structure of soft-loud-soft-loud-soft-loud. By comparing the behavioral and neural data in these four conditions, we aimed to uncover the impact of intensity arrangement on task performance and neural correlates, focusing particularly on the phase information in EEG data. We also ran the condition where the sound intensities were equal throughout the trials.

The behavioral data show that listeners scored over 80 % on average when they attended to louder targets (80 dB), whereas they performed near chance when the target sounds were softer (30 dB), demonstrating level dominance. However, it was unexpected that the performance when the sound intensities of odd and even components were equal (55 dB) was also near chance and even slightly poorer than the Attend Soft condition. Inconsistency with the past study (Laffere et al., 2020; Laffere et al., 2021) may be due to how much frequency overlap exists between targets and nontargets. They used a pattern of notes alternating between low and high frequencies that are an octave apart, which made two sound streams non-overlapping in both time and frequency. In contrast, in the current study, two sets of sounds occupied the same spectral region (bandpass filtered at the same cutoff frequencies). This left only the temporal non-overlap as cues for selective attention, making the task overall harder than the one used in Laffere et al. (2020). The only way to achieve higher accuracy was to use intensity difference as a supporting cue (as in Attend Loud condition). Additionally, the fact that six out of sixteen listeners performed better in the Attend Soft condition compared to the Attend Equal condition may be attributed to the degree of target-nontarget similarity. In the Attend Soft condition, despite the detrimental effect of louder nontargets, there is a 50-dB difference between targets and nontargets, which could have helped those listeners perform better than in the Attend Equal condition, where no such cue was present. Past studies have shown that target-masker similarity influences performance in several behavioral tasks (Kidd Jr et al., 1994; Durlach et al., 2003; Lee and Richards, 2011), and that intensity difference can serve as a cue for, although not as robust as frequency cues (Rose and Moore, 2000). However, their performance was still far from the accuracy achieved in the Attend Loud condition.

To interpret our EEG data, we constructed polar plots providing a visual presentation of the average neural phase angles and inter-trial phase coherence (ITPC) computed based on trial-by-trial phase information. Contrary to our expectation, the phase shift data did not provide clear insights into listeners' attentional direction under four conditions. The first set of polar plots used the *Attend Odd: 80 dB* condition as a baseline and shows the individual average neural phase angles and ITPC in the other conditions relative to this baseline. The polar plot for *Attend Even: 30 dB* shows that the majority of phase angles in this condition were biased towards the right side of the polar coordinate, suggesting that listeners' direction was not directed to the target-relevant temporal position but rather towards the locations of the louder components. Also, despite variations in behavioral data, no association between task performance and neural phase angle was demonstrated. Similar trends were observed when *Attend Even: 80 dB* was used as a baseline.

In conclusion, our study failed to observe an EEG signal indicating successful selective attention to lower-level targets. Instead, listeners' attention seemed to be automatically directed towards the higher-level nontargets, demonstrating the robustness of level dominance. This finding aligns with the arguments made by previous studies (e.g., Oberfeld et al., 2013; Turner and Berg, 2007), emphasizing the difficulty in ignoring louder components even when they are irrelevant to the task.

Our results suggest that involuntary attention capture by louder stimuli can override voluntary attention mechanisms. Still, to deeply investigate the interplay of the two types of attentional mechanisms in the context of level dominance, additional studies involving different neural metrics or experimental designs may be required.

CONCLUSION

The experiments reported in this dissertation address three questions concerning auditory level dominance.

Chapter 1 explored whether changes in sound promoting sound stream segregation also reduce level dominance. We studied how variations in sound qualities affect level dominance using a multi-sound intensity discrimination task. We focused on three types of acoustic cues known to promote compulsory stream segregation: changes in pitch, timbre, and frequency region. The findings show that alterations in pitch or timbre do not lead to an effective reduction in level dominance; however, reducing frequency overlap between sounds significantly reduces this effect. This suggests that difference in frequency extent, a potent acoustic cue for stream segregation, is also effective in reducing level dominance.

Chapter 2 examined the relation between level dominance and gap duration discrimination tasks (as a proxy for stream segregation) through various sound manipulations. Both individual and average correlations reveal a strong link between the two tasks. We found that the degree of stream segregation is systematically related to the release from level dominance observed in the same subjects, suggesting that they may share common underlying mechanisms.

Chapter 3 transitioned to the electrophysiological domain, investigating how the relative intensity of sequential sounds impacts selective attention. In the pitch discrimination task, listeners were instructed to attend to the odd-numbered components in the sound sequence for half of the session and to the even-numbered components for the other half. We aimed to explore the effects of sound levels on selective attention both behaviorally and electrophysiologically. Behavioral results showed that accuracy was high when listeners attended to high-level targets but declined to near chance when the targets were at the same or lower sound levels than the non-targets. Electrophysiological data from EEG recordings indicated that listeners' attention was directed toward louder components even when they were instructed to attend to the softer components. Our results suggest that involuntary attention capture by louder sounds can override voluntary attention mechanisms.

Bibliography

Abel, S. M. (1972). Discrimination of temporal gaps. *The Journal of the Acoustical Society of America*, 52(2B):519–524.

Beauvois, M. W. and Meddis, R. (1996). Computer simulation of auditory stream segregation in alternating-tone sequences. *The Journal of the Acoustical Society of America*, 99(4):2270–2280.

Berens, P. (2009). Circstat: a matlab toolbox for circular statistics. *Journal of statistical software*, 31:1–21.

Berg, B. G. (1990). Observer efficiency and weights in a multiple observation task. *The Journal of the Acoustical Society of America*, 88(1):149–158.

Brainard, D. H. and Vision, S. (1997). The psychophysics toolbox. *Spatial vision*, 10(4):433–436.

Bregman, A. S., Liao, C., and Levitan, R. (1990). Auditory grouping based on fundamental frequency and formant peak frequency. *Canadian Journal of Psychology/Revue canadienne de psychologie*, 44(3):400.

Chen, X., Xu, B., Chen, Y., Zeng, X., Zhang, Y., and Fu, S. (2023). Saliency affects attentional capture and suppression of abrupt-onset and color singleton distractors: Evidence from event-related potential studies. *Psychophysiology*, 60(8):e14290.

Corey, D. M., Dunlap, W. P., and Burke, M. J. (1998). Averaging correlations: Expected values and bias in combined pearson rs and fisher's z transformations. *The Journal of General Psychology*, 125(3):245–261.

Cusack, R. and Roberts, B. (2000). Effects of differences in timbre on sequential grouping. *Perception & psychophysics*, 62:1112–1120.

Dai, L. and Shinn-Cunningham, B. G. (2016). Contributions of sensory coding and attentional control to individual differences in performance in spatial auditory selective attention tasks. *Frontiers in Human Neuroscience*, 10:530.

Delorme, A. and Makeig, S. (2004). Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. *Journal of neuroscience methods*, 134(1):9–21.

Ding, N. and Simon, J. Z. (2012). Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. *Journal of neurophysiology*, 107(1):78–89.

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

Hartmann, W. M. and Johnson, D. (1991). Stream segregation and peripheral channeling. *Music perception*, 9(2):155–183.

Hyvärinen, A. and Oja, E. (1997). A fast fixed-point algorithm for independent component analysis. *Neural computation*, 9(7):1483–1492.

Iverson, P. (1995). Auditory stream segregation by musical timbre: effects of static and dynamic acoustic attributes. *Journal of Experimental Psychology: Human Perception and Performance*, 21(4):751.

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

Laffere, A., Dick, F., Holt, L. L., and Tierney, A. (2021). Attentional modulation of neural entrainment to sound streams in children with and without adhd. *NeuroImage*, 224:117396.

Laffere, A., Dick, F., and Tierney, A. (2020). Effects of auditory selective attention on neural phase: individual differences and short-term training. *NeuroImage*, 213:116717.

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

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

Lutfi, R. A. and Jesteadt, W. (2006). Molecular analysis of the effect of relative tone level on multitone pattern discrimination. *The Journal of the Acoustical Society of America*, 120(6):3853–3860.

Madsen, S. M., Whiteford, K. L., and Oxenham, A. J. (2017). Musicians do not benefit from differences in fundamental frequency when listening to speech in competing speech backgrounds. *Scientific Reports*, 7(1):12624.

Micheyl, C. and Oxenham, A. J. (2010). Objective and subjective psychophysical measures of auditory stream integration and segregation. *Journal of the Association for Research in Otolaryngology*, 11:709–724.

Moore, B. C. and Gockel, H. (2002). Factors influencing sequential stream segregation. *Acta Acustica United with Acustica*, 88(3):320–333.

Moore, B. C. and Gockel, H. E. (2012). Properties of auditory stream formation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1591):919–931.

Oberfeld, D. (2008). The mid-difference hump in forward-masked intensity discrimination. *The Journal of the Acoustical Society of America*, 123(3):1571–1581.

Oberfeld, D., Kuta, M., and Jesteadt, W. (2013). Factors limiting performance in a multitone intensity-discrimination task: Disentangling non-optimal decision weights and increased internal noise. *PloS one*, 8(11):e79830.

Oberfeld, D. and Stahn, P. (2012). Sequential grouping modulates the effect of non-simultaneous masking on auditory intensity resolution. *PloS one*, 7(10):e48054.

Oberfeld, D., Stahn, P., and Kuta, M. (2012). Binaural release from masking in forward-masked intensity discrimination: Evidence for effects of selective attention. *Hearing Research*, 294(1-2):1–9.

Oh, Y., Zuwala, J. C., Salvagno, C. M., and Tilbrook, G. A. (2022). The impact of pitch and timbre cues on auditory grouping and stream segregation. *Frontiers in Neuroscience*, 15:725093.

Ponsot, E., Susini, P., Saint Pierre, G., and Meunier, S. (2013). Temporal loudness weights for sounds with increasing and decreasing intensity profiles. *The Journal of the Acoustical Society of America*, 134(4):EL321–EL326.

Relkin, E. M. and Doucet, J. R. (1991). Recovery from prior stimulation. i: Relationship to spontaneous firing rates of primary auditory neurons. *Hearing research*, 55(2):215–222.

Remington, R. W., Johnston, J. C., and Yantis, S. (1992). Involuntary attentional capture by abrupt onsets. *Perception & Psychophysics*, 51(3):279–290.

Richards, V. M., Shen, Y., and Chubb, C. (2013). Level dominance for the detection of changes in level distribution in sound streams. *The Journal of the Acoustical Society of America*, 134(2):EL237–EL243.

Rose, M. M. and Moore, B. C. (2000). Effects of frequency and level on auditory stream segregation. *The Journal of the Acoustical Society of America*, 108(3):1209–1214.

Ruggles, D. R., Freyman, R. L., and Oxenham, A. J. (2014). Influence of musical training on understanding voiced and whispered speech in noise. *PloS one*, 9(1):e86980.

Samaha, J. and Postle, B. R. (2017). Correlated individual differences suggest a common mechanism underlying metacognition in visual perception and visual short-term memory. *Proceedings of the Royal Society B: Biological Sciences*, 284(1867):20172035.

Schlauch, R. S., Lanthier, N., and Neve, J. (1997). Forward-masked intensity discrimination: Duration effects and spectral effects. *The Journal of the Acoustical Society of America*, 102(1):461–467.

Schröger, E. (1996). A neural mechanism for involuntary attention shifts to changes in auditory stimulation. *Journal of cognitive neuroscience*, 8(6):527–539.

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

Silver, N. C. and Dunlap, W. P. (1987). Averaging correlation coefficients: Should fisher's z transformation be used? *Journal of applied psychology*, 72(1):146.

Srinivasan, R. and Petrovic, S. (2006). Meg phase follows conscious perception during binocular rivalry induced by visual stream segregation. *Cerebral Cortex*, 16(5):597–608.

Summers, R. J., Bailey, P. J., and Roberts, B. (2010). Effects of differences in fundamental frequency on across-formant grouping in speech perception. *The Journal of the Acoustical Society of America*, 128(6):3667–3677.

Tan, A. Y. and Berg, B. G. (2018). Level dominance effect and selective attention in a dichotic sample discrimination task. *The Journal of the Acoustical Society of America*, 143(4):2119–2127.

Turner, M. D. and Berg, B. G. (2007). Temporal limits of level dominance in a sample discrimination task. *The Journal of the Acoustical Society of America*, 121(4):1848–1851.

van Noorden, L. P. A. S. (1975). Temporal coherence in the perception of tone sequences.

Vliegen, J. and Oxenham, A. J. (1999). Sequential stream segregation in the absence of spectral cues. *The Journal of the Acoustical Society of America*, 105(1):339–346.

Watanabe, K., Srinivasan, R., and Richards, V. M. (2023). Effects of sound segregation cues on multi-sound intensity discrimination. *JASA Express Letters*, 3(12).

Zar, J. H. (1999). *Biostatistical analysis*. Pearson Education India.

Zeng, F.-G. and Turner, C. W. (1992). Intensity discrimination in forward masking. *The Journal of the Acoustical Society of America*, 92(2):782–787.

Zeng, F.-G., Turner, C. W., and Relkin, E. M. (1991). Recovery from prior stimulation ii: Effects upon intensity discrimination. *Hearing research*, 55(2):223–230.

Appendix A

A.1 Computing idealized local boundary

Note that an alternative strategy to solve the temporal-gap duration discrimination task in Chapter 2 is to focus on the last triplet of a stimulus sequence (refer to Figure 2.1), which consists of nontarget, target, and another nontarget, and perform a local comparison to adjacent sounds to judge the direction of the temporal shift in the last target sound. In other words, listeners could compare the duration of the gap between the first nontarget and target (G_1), and the gap between the target and second nontarget (G_2) to make a decision. The decision rule is: Press "1" if $G_1 < G_2$, and press "2" otherwise (Figure A.1).

To demonstrate that this local comparison is insufficient as the ideal cue for this task, we performed a computer simulation to compute the idealized local boundary B of Δt where a noiseless observer can achieve 71% correct (Levitt, 1971) just by comparing the last target to two adjacent nontargets. The simulation computes $B \approx 26$ ms, indicating that for $\Delta t \geq 26$ ms, local comparison cue may be contributing to the performance, but the conditions with threshold $\Delta t < 26$ ms (*lower f_o different region, noise and off-freq*) indicate that local comparison is not the sole cue and some additional cue (presumably within-target

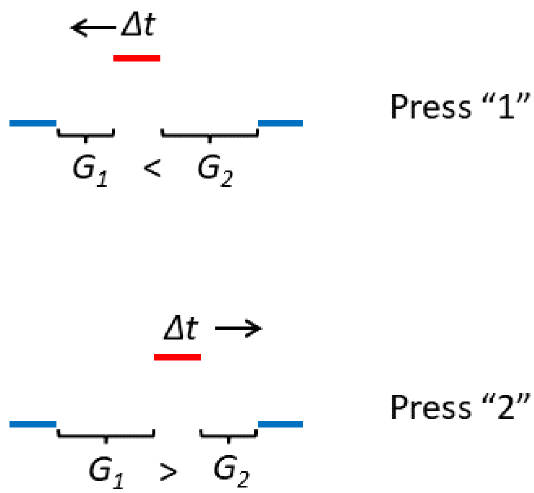

Figure A.1: Schematic of how a noiseless observer would focus on the last triplet of a stimulus sequence and perform a local comparison to solve the temporal-gap duration discrimination task in Chapter 2. In the top case, the target (red) is shifted to the left, causing the gap between the first nontarget (blue) and target, G_1, to be shorter than the gap between the target and the second nontarget G_2. The ideal observer would press "1" in this case, and press "2" otherwise (bottom).

comparison) is incorporated to achieve those threshold values.