THE UNIVERSITY OF CHICAGO

FLEXIBLE CHANGES IN NEURAL GAIN AND SELECTIVITY IN THE EARLY AUDITORY ENCODING OF EXPERTS UNDER VARYING ATTENTION AND EFFORT DEMANDS

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE SOCIAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF PSYCHOLOGY

BY
LETITIA HO

CHICAGO, ILLINOIS
AUGUST 2025

ABSTRACT

The frequency-following response (FFR) offers a powerful window into how attention and domain-specific expertise shape early auditory encoding. Yet, despite decades of research, the cognitive role of the FFR remains underdefined. Existing studies of attention often report mixed or null results, suggesting that its influence on early auditory encoding may be weak or contingent on individual factors such as expertise. Conversely, studies of expertise consistently show enhanced FFR amplitudes in musicians and other experts (e.g., Wong et al., 2007; Parbery-Clark et al., 2009; Strait et al., 2013), but these have largely been conducted under passive listening conditions, without explicit attentional demands. As a result, the literature defaults to the assumption that larger FFR amplitudes signify superior encoding, yet this assumption lacks strong theoretical grounding. To address this gap, the present dissertation examines the joint effects of attention and expertise across active and passive listening conditions, drawing on competing predictions from gain modulation models (e.g., Hillyard et al., 1998; Zatorre et al., 1999, 2012) and dynamic tuning accounts (Fritz et al., 2003; Fritz et al., 2007; Elhilali et al., 2007). Across four experiments, the findings reveal that expertise sharpens FFR responses in a non–context-specific manner, consistent with dynamic tuning models that emphasize selective suppression of irrelevant neural activity. However, under more challenging tasks, effort in experts was associated with global increases in FFR amplitude, suggesting the flexible recruitment of gain-modulation mechanisms to strengthen performance. Passive listening, by contrast, consistently elicited broadly elevated FFR amplitudes, indicating a diffuse, non-selective form of auditory engagement. Taken together, these results suggest that early auditory encoding flexibly engages mechanisms of both selectivity and gain, depending on task difficulty, effort, and expertise. The findings contribute toward a more comprehensive model of auditory attention in which dynamic tuning enhances encoding efficiency under favorable conditions, while gain modulation amplifies responses to meet the demands of effortful listening.

This dissertation is dedicated to my parents, thank you both for being so wonderful. My sincerest apologies for never telling you about anything I'm doing until after I'm finished.

CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

This dissertation would not have been possible without the support of so many people in my life. My biggest thanks first and foremost to Howard, for your steady mentorship. As I tell you, and joke to my friends, you are the biggest big-picture guy and also a grant-writing wizard. You manage to communicate the most emotions, from excitement to disappointment, in the shortest, most laconic emails. You have the best all-knowing smile, even among the wizened faculty here at the University of Chicago. Thank you also to the rest of my committee, for your intellectual guidance and practical advice. I also want to thank my mom, my dad, my brothers, my dog and my late cats. Dad, you manage to both be the silliest person in my life, and somehow maintain more integrity than anyone else. Mom, sorry for stressing you out all the time, but thank you for worrying about me. Lewis, thanks for shouldering most of our parents' anxieties. Levin, thanks for being so good. Thank you Jack, for keeping me company and broadening my intellectual horizons. Thank you Ezra, for always looking out for me. Thanks, Carolyn, for seeing through my and everyone else's pretenses all the time. Grant and Layn, you are the best people I know. Thank you to my frisbee friends and teammates for running around outside with me. To my friends who are not in Chicago, thanks for being a part of this journey from afar. I hope you know that I miss you. Lisa, you know me like no one else. Mona, we are almost too similar in some ways, but for that reason I always love talking to you. Thanks to my library friends for keeping me sane. And last but not least, thank you Tommy for always making me smile.

CHAPTER 1

INTRODUCTION

The frequency-following response (FFR) offers a powerful tool for investigating how attention and domain-specific expertise shape early auditory encoding. The existing FFR literature presents a complex and sometimes contradictory picture. Studies of attentional effects often report mixed or null results, suggesting that attention's influence on early auditory encoding may be weak or contingent on individual differences such as expertise. On the other hand, a separate body of work examining expertise has consistently shown enhanced FFR amplitudes in musicians and other experts (e.g., Wong et al., 2007; Parbery-Clark et al., 2009; Strait et al., 2013). However, nearly all such expertise studies have been conducted under passive listening conditions, where participants are not given explicit task demands or attentional goals. Thus, the literature on attention effects tends to overlook expertise, while the literature on expertise effects has largely ignored the role of active attentional engagement. These gaps have resulted in a literature that remains sparse in terms of a cognitive model of what the FFR actually signifies. Most studies implicitly assume that larger FFR amplitudes reflect better auditory encoding, yet this assumption has rarely been tested or theoretically justified. There remains a need for a framework linking amplitude and phase consistency to specific attentional or expertise-driven mechanisms to establish a clear cognitive model of what the FFR represents. Two dominant classes of theoretical models make divergent predictions about the role of attention and expertise in later, higher-level auditory processes. Given the underspecified nature of cognitive models of the FFR, these models may be borrowed to make predictions about the effects of attention and expertise on the FFR. Gain modulation models are grounded in evidence from human neuroimaging and electrophysiology, which measure low-frequency, large-scale population responses such as the BOLD signal in fMRI or late event-related potentials (ERPs). These studies consistently show that attention and expertise enhance neural activity by increasing response gain in relevant sensory cortices (e.g., Hillyard et al., 1998; Zatorre et al., 1999, 2012). From this perspective, experts under task demands should exhibit stronger FFR amplitudes, reflecting amplified neural responses to relevant features. In contrast, dynamic tuning models are informed by single- and multi-unit recordings from animal primary auditory cortex. These studies demonstrate that attention and expertise increase the selectivity of neural firing, enhancing responses to relevant features while suppressing activity to irrelevant or distracting inputs (Fritz et al., 2003; Fritz et al., 2007; Elhilali et al., 2007). Taken together, these frameworks suggest competing accounts of how early auditory encoding is modulated. Studies on late, low-frequency responses suggest that the brain amplifies neural responses, whereas accounts based on firing rates of neurons in the cortex suggest responses are sharpened through selective suppression. By borrowing predictions from models informed by later auditory processes, this work seeks to adjudicate between amplification and selectivity accounts, and to test the conditions under which their principles extend to the earliest stages of auditory encoding. In short, this dissertation studies the interactions between task demands and expertise to investigate how attention and long-term experience influence early auditory processing to build a model of the cognitive role of the FFR in audition and auditory attention.

There is a large body of existing literature on the attention effects on the FFR, but its results are famously mixed. Intermodal studies asking participants to focus on either a visual stimuli or auditory stimuli stream have found that selective attention modifies FFR amplitude (Galbraith et al., 2003; Hairston et al., 2013; Lehmann & Schönwiesner, 2014), while other studies have not (Galbraith & Kane, 1993; Galbraith & Doan, 1995). Results from dichotic listening studies are similarly inconsistent (Galbraith & Arroyo, 1993; Galbraith et al., 1998; Lehmann & Schönwiesner, 2014). Hoorman and colleagues (2000) and Varghese and colleagues (2015) both tested attention effects on FFR amplitudes and latency using both intermodal and dichotic tasks. The former found only moderate latency effects in the fourth experiment using stimulus pairs, while the latter found no attention effects on phase locking. Recent studies on the attentional modulation of the early auditory response suggest that top-down attentional effects may be limited to cortical responses. An MEG study by Hartmann and Weisz (2019) found that only activity in the right PAC was affected by shifts in intermodal attention. As the cortical component of the FFR is limited to low frequencies (<200 Hz), attentional modulation may be limited to low-frequency stimulus features (Bidelman, 2018; Holmes et al., 2018).

The mixed results of top-down attention on the FFR suggest that modulations of selective attention induced by task demands create changes in early auditory processing that are too fine-grained to reliably detect. One potential confound is the success of the selected attention manipulations. Past studies have shown the importance of analyzing task performance to assess whether task manipulations effectively drive attention (Goh et al., 2014; Engelmann et al., 2009). Attention effects may be stronger in participants who successfully deploy selective attention to monitor certain frequency bands or stimulus features. To our knowledge, there are no existing studies that compare the interaction between expertise and task demands on the FFR. This may be partly attributed to the low sample sizes historically used in studies on attention effects in the FFR (e.g., Galbraith et al., 2003; Hoorman et al., 2004; Lehmann & Schönwiesner, 2014; Price & Bidelman, 2021; Schuller et al., 2023; Bharadwaj et al., 2014). Investigating the interactions between task demands and task performance may provide more insight into how attention interacts with prior experience to influence the FFR, and may explain the mixed results in prior literature.

There is an existing, parallel body of literature on the effects of expertise on the FFR. These studies on the FFRs in musicians or bilinguals typically report increases in FFR amplitude and consistency with expertise, however FFRs signals reported in these studies are recorded exclusively under passive listening conditions. For example, studies have found enhanced FFRs to auditory stimuli in musicians (Parbery-Clark et al., 2009; Strait et al., 2009; 2013; Wong et al., 2007). Parbery-Clark and colleagues (2009) compared FFRs recorded passively to speech and speech in noise in musicians and non-musicians. They found that musicians had lower latency FFRs, greater power at the stimulus harmonics, and higher correlations between their FFRs and the stimulus signal, and between the FFRs to speech and speech in noise. Similarly, Musacchia and colleagues (2008) report both early latency and higher amplitude FFRs in musicians compared to non-musicians to different speech and music tokens. Comparable advantages in early auditory responses have been recorded in tonal language speakers (Jeng et al., 2011; Krishnan et al., 2005) and bilinguals (Krizman et al., 2014, 2015, 2016; Skoe et al., 2017; Omote et al., 2017). These studies may have separated behavioral tasks from electrophysiological recordings due to the high number of trials needed to record FFR signals (Krizman & Kraus, 2010; Bidelman, 2018). Even studies that looked at the effects of training on early auditory encoding recorded FFRs separately from the tasks used to measure learning and perceptual improvement (Russo et al., 2005; Song et al., 2008).

While the literature on the effects of expertise on early auditory encoding have led to the assumption that greater FFR amplitudes are indicative of enhanced early auditory encoding, cognitive models of perceptual processing such as predictive coding models (Rao & Ballard, 1999; Friston, 2005) and biased competition models (Chelazzi, 1993; Li et at., 1993) predict a more complicated relationship between neural gain and encoding fidelity. Predictive coding accounts propose that the brain continuously generates top-down predictions about incoming sensory input and computes the difference between these predictions and the actual input to guide perception and learning (Rao & Ballard, 1999; Friston, 2005). According to this framework, neural activity in sensory regions largely reflects prediction error rather than raw input, so when a stimulus is expected or familiar, the top-down prediction closely matches the input, leading to reduced neural responses. This stands in contrast to gain modulation theories, which posit increased sensory activity for attended or behaviorally relevant stimuli. In predictive coding, expertise leads to more accurate internal models of the sensory environment, thus minimizing prediction error and consequently reducing sensory cortex activation (Auksztulewicz & Friston, 2016; Blank & Davis, 2016). These models account for findings where repetition, familiarity, or predictability result in suppressed activity in auditory or visual cortices, challenging the notion that increased neural response always indicates better perceptual processing (Summerfield & Egner, 2009; Todorovic et al., 2011). Predictive coding therefore suggesting that attenuated activity can reflect more efficient processing.

Similarly, biased competition models, originally developed in the context of visual processing, propose that multiple sensory inputs compete for limited neural resources and that attention biases this competition in favor of task-relevant or behaviorally salient stimuli (Desimone & Duncan, 1995). Top-down attention biases this competition such that attended stimuli are more likely to win, resulting in a larger population of higher-order neurons encoding the attended stimulus and implicitly binding its features together (Chelazzi et al., 1993; Desimone & Duncan, 1995; Reynolds & Desimone, 1999). During the delayed match-to-sample task, Chelazzi (1993) and colleagues demonstrated that activity in neurons tuned to a particular feature was reduced if the stimulus with the preferred feature was not the target. Moreover, intracranial recordings during selective visual attention reveal a shrinkage in the population of activated neurons as stimuli become more familiar and responses from cells not tuned to the properties of the attended stimulus are suppressed. Li and colleagues (1993) found that responses to familiar visual stimuli decreased firing in most units, with responses in one-third of recorded cells decreasing as stimuli became more familiar. These results suggest that familiarity and expertise sharpen neural selectivity by enhancing the responsiveness of neurons tuned to relevant features while suppressing activity to irrelevant stimuli. Crucially, the increase in selectivity does not necessarily result in a net increase in the overall amplitude of neural activity.

Given the mixed results of the existing literature on the effects of top-down attention on the FFR, and the blind spot in the current literature on the effects of expertise and long-term experience on the FFR, it is important to study how experience and attention interact to modulate early auditory encoding. Attention does not refer to a specific neurobiological mechanism, rather it is an umbrella term for a complex jumble of interacting processes that determine the effects of immediate task demands, goals, and context on mental processes. Existing FFR literature rarely provides an explicit cognitive model about the role of the FFR in audition and auditory attention. Instead, based on the findings of greater FFR amplitudes in expertise during passive listening, these models typically assume that greater FFR amplitudes correspond to better or enhanced stimulus encoding. Cognitive models of perceptual processing such as predictive coding and biased competition, however, suggest that opposite patterns of response amplitudes may be observed in experts under task demands. Due to the aforementioned gaps in the literature on top-down attention and expertise effects on the FFR, predictions about the effects of attention and expertise and their interactions on the FFR will be borrowed from a robust literature on the effects of attention and expertise on later, downstream, parts of auditory processing. These models will be split roughly into two classes, first are gain modulation models, based mostly on neuroimaging data reflecting late, low-frequency, large population responses, that suggest an overall increase in neural activity with attention or expertise. Second are groups of models that will be referred to as dynamic tuning models, which find primarily an increase in the selectivity in the firing rates of individual neurons in primary sensory cortices in response to attention and expertise. The pervasiveness of gain modulation models, a class of models that share the same assumption about increases in neural activity corresponding to enhanced encoding, may have informed the assumption that early auditory encoding, reflected by the FFR, may follow similar principles.

Gain modulation theories occupy a foundational and influential place within the broader canon of theories about perceptual attention and neural representation. Gain modulation models posit that attention enhances neural responses to relevant sensory input by increasing the amplitude of neural responses to attended or behaviorally salient stimuli (Churchland, 2016; Salinas & Thier, 2000). These models are notable for attempting to bridge cognitive models, such as Posner's spotlight model (1980), which described attention as enhancing perceptual processing at a given location or feature, with neurobiological data. While early evidence for gain modulation were intracranial recordings in animals showing that visual attention increases the gain of neurons in the visual cortex (Posner and Gilbert 1999; McAdams & Maunsell, 1999; Treue & Martinez Trujjillo, 1999), gain-like increases in the amplitude of neural activity have since been found using neuroimaging monitoring BOLD activity (Kastner & Ungerleider, 2001; O'Craven et al., 1999; Gauthier et al., 2000) and ERPs (Hillyard et al., 1973; Näätänen et al., 1978) and neuroimaging studies in humans have been relied on as some of the strongest, most foundational evidence for gain modulation theories. Contemporary gain modulation accounts of perceptual processing in humans inherit from this tradition and rely similarly heavily on fMRI data.

Gain modulation theories have been applied across modalities to explain auditory attention. A review by Zatorre, Belin, and Penhune (2002) presented evidence in which participants displayed increased BOLD activity in bilateral auditory cortices when listening to complex sounds, suggesting that attentional and training-related mechanisms modulate neural responses. Zatorre and colleagues (2012) argue that experience amplifies the responses of relevant neural populations, a hallmark of gain modulation. This aligns with early MEG work by Pantev and colleagues (1998), which found that musicians exhibit significantly larger auditory-evoked fields in response to piano tones compared to non-musicians, particularly in the primary auditory cortex. Similarly, Golestani and Zatorre (2004) investigated phonetic learning and its neural correlates, revealing that individuals who acquired non-native phonetic contrasts more successfully elicited greater structural and functional activity in the left auditory cortex. Across these studies, the pattern of results consistently indicates that the auditory system, like the visual system, can undergo gain-like changes in response to different top-down demands. These increases in gain are posited to amplify the neural representation of attended stimuli.

More recently, EEG and MEG data have been used to build accounts of attentional gain modulation at higher frequencies of neural activity. While models of auditory attention informed by fMRI data posit the neurobiological mechanism of attention as shifts in relative activity in domain-specific regions on a second-based timescale, EEG has allowed the measurement of fluctuations in high-frequency oscillations in the brain, informing models of attention such as dynamic attending theory. Large and Jones (1999) argue that selective sensory attention is instantiated by the entrainment of endogenous, ongoing neural oscillations to external rhythms. This entertainment is proposed to align periods of increased excitability in neural populations with behaviorally salient external events (Lakatos et al., 2019; Obleser & Kayser, 2019). According to entrainment models, synchronizing internal oscillatory activity, such as delta, theta, and gamma waves, to rhythmic stimuli such as music and speech stimuli can amplify the neural representation of stimuli that occur at periods of high gain, enhancing the signal-to-noise ratio of neural encoding (Doelling & Poeppel, 2015; Luo & Poeppel, 2007; Ding & Simon, 2014). A study on macaques by Lakatos and colleagues (2008) using intracranial recordings found that attending to rhythmic auditory or visual streams leads to neural oscillations in the auditory cortex that align in phase with the rhythm, thereby boosting sensory responses when stimuli are expected. Similar results have been found using auditory stimuli. Gross and colleagues (2013) recorded MEG in participants during continuous speech listening and found that delta and theta oscillations in the auditory cortex were coupled to the temporal structure of speech, while gamma-band power nested within these rhythms tracked fine acoustic details. Luo and Poeppel (2007) report that theta band (4-8 Hz) activity reliably tracks syllable-level information in spoken sentences, with increased classification performance correlating with the intelligibility of the sentence.

It is important to note that these increases in neural gain with task demands or expertise have generally been observed from late, low-frequency responses measured by blood oxygenation level dependent (BOLD), event related potentials (ERPs) or fluctuations in theta- and delta-band activity. The late responses, typically aggregated over large populations of neurons, may reflect late, decision making or integrative processes. While gain like changes in activity may occur in these late processes with attention or changes in experience, this general principle of increased neural activity corresponding to enhanced perceptual processing may not apply to a response that measures an early, high-frequency response such as the FFR. The pervasiveness of these findings has implicitly bolstered the assumption that greater FFR amplitudes indicate enhanced auditory encoding. Other theories of perceptual processing such as biased competition models and dynamic tuning, based often on single- or multi-unit electrophysiological recordings in animals, argue for an increase in the selectivity of neural firing, instead of large-amplitude or population-wide increases in activity or gain.

In contrast to gain modulation models, biased competition models, originally developed in the context of visual processing, propose that multiple sensory inputs compete for limited neural resources and that attention biases this competition in favor of task-relevant or behaviorally salient stimuli (Desimone & Duncan, 1995). Top-down attention biases this competition such that attended stimuli are more likely to win, resulting in a larger population of higher-order neurons encoding the attended stimulus and implicitly binding its features together (Chelazzi et al., 1993; Desimone & Duncan, 1995; Reynolds & Desimone, 1999). During the delayed match-to-sample task, Chelazzi (1993) and colleagues demonstrated that activity in neurons tuned to a particular feature was reduced if the stimulus with the preferred feature was not the target. Moreover, intracranial recordings during selective visual attention reveal a shrinkage in the population of activated neurons as stimuli become more familiar and responses from cells not tuned to the properties of the attended stimulus are suppressed. Li and colleagues (1993) found that responses to familiar visual stimuli decreased firing in most units, with responses in one-third of recorded cells decreasing as stimuli became more familiar. These results suggest that familiarity and expertise sharpen neural selectivity by enhancing the responsiveness of neurons tuned to relevant features while suppressing activity to irrelevant stimuli. Crucially, the increase in selectivity does not necessarily result in a net increase in the overall amplitude of neural activity.

Dynamic tuning models extend certain principles from biased competition models in vision to understanding how auditory attention shapes neural processing at the level of single neurons and neural populations, particularly in the primary auditory cortex. Research by Fritz and colleagues (2003, 2007) has shown that neurons in the primary auditory cortex dynamically change their spectrotemporal receptive fields (STRFs) depending on behavioral context. In their studies on ferrets, directing an animal's attention to a behaviorally relevant sound feature, such as a tone embedded in noise, led to rapid, reversible changes in STRFs. The dimension-selective model by Holt and colleagues (2018) reviewed a study by Schwartz and David (2018) in which mice were presented with two simultaneous and spatially separated streams of tones containing both distractor tones and target tones. They report that spike rates of most primary auditory cortex units decreased to distractor tones in the narrowband noise around the target tone, whereas spike rates to the target tone were not influenced by whether its frequency band was attended or ignored. Such plasticity suggests that auditory attention selectively enhances the firing of neurons tuned to task-relevant features by modifying receptive field properties in real time and suppresses the responses of non-feature-selective neurons. These models, extended to the FFR, predict an increase in the selectivity of the early auditory response in experts under task demands. Dynamic tuning therefore predicts a reduction in the overall amplitude of the response, contrary to the assumptions of the current literature.

Different models of auditory attention have emerged in part due to the particular neuroimaging and electrophysiological methods that inform them. Models grounded in human fMRI, such as gain modulation theories, focus on changes in activity in regions of interest, characterizing auditory attention in terms of enhanced BOLD responses in broad primary and non-primary auditory cortices and frontotemporal networks (Alho et al., 2003; Hill & Miller, 2010; Paltoglou et al., 2009; Holt et al., 2018; Luthra et al., 2024). These data typically reflect the activity of late, integrative or decision-making related processes. In contrast, animal electrophysiology studies, which index the firing of individual neurons at a high temporal resolution, support models like biased competition and dynamic tuning, which emphasize attentional sharpening of receptive fields in primary auditory cortex (Fritz et al., 2003; Fritz et al., 2007; David et al., 2012). Still other frameworks, grounded in EEG and MEG, conceptualize attention as a dynamic rhythmic sampling process, in which neural oscillations entrain to the temporal dynamics of attended sounds (Large & Jones, 1999; Lakatos et al., 2008; Schroeder & Lakatos, 2009; Obleser & Kayser, 2019). These methodological differences inform that predictions that these models might make on the effects of attention and expertise on the FFR.

The FFR offers a promising intermediate for reconciling these divergent theoretical frameworks and the methods that support them. Unlike fMRI, which provides coarse spatial resolution, or single-unit recordings, which offer data from isolated neurons, the FFR captures population-level, phase locked activity in subcortical and cortical auditory structures with high temporal precision. Evidence of the short latency (6-8 ms) of the FFR suggests that it plays a role in early auditory encoding (Galbraith et al., 1990; Hoormann et al., 1994; King et al., 2016). Despite this, studies have shown that FFR amplitude and fidelity are modulated by attentional state and auditory expertise (Hartmann & Weisz, 2019; Bidelman & Alain, 2015; Coffey et al., 2017, 2019), suggesting that attention enhances neural synchrony without necessarily increasing average firing rates. This aligns FFR data more closely with the tuning and sparsity effects observed in animal electrophysiology, while also reflecting population-level changes observed in fMRI but at a higher frequency. As such, the FFR may help bridge the gap between hemodynamic signals and the firing patterns of individual neurons, offering a more integrated view of how selective attention shapes auditory processing across measurement modalities.

The studies in this dissertation will compare measures of FFR amplitude with measures of phase-locking consistency. While studies in the FFR literature typically use one measure or the other and take both to generally reflect FFR 'strength' (Omote et al., 2017; Krizman et al., 2016; Varghese et al., 2015; Bharadwaj et al., 2015), the two variables are complementary but may vary independently. Power or amplitude reflects the magnitude of a neural response at a given frequency, and as with other aggregate electrophysiological responses, is typically regarded as a measure of the size of the neural population engaged in the response. Greater FFR amplitudes may suggest phase-locked neural firing from a bigger population of neurons. However, FFR amplitude may also vary if the consistency of phase locking reduces due to wave interference. It is therefore imperative to compare measures of FFR amplitude with changes in FFR consistency to determine whether the differences in observed response amplitude should be attributed to changes in the size of the population engaged in the synchronized response or to changes in the temporal precision of the response. Here, phase locking consistency is measured by the phase locking value (PLV) (Lachaux et al., 1999; Zhu et al., 2013), which is computed by extracting phase information at the frequency of interest from the Fourier transform of each trial. The phase angle of each trial is represented by a unit-length complex exponential. These unit vectors are then averaged across trials, with the magnitude of the resulting vector, the PLV, indexing the consistency of phase alignment across trials. To assess significance, observed PLVs are compared against a null distribution generated by randomizing phase values across trials. This approach provides a robust measure of temporal synchronization that is independent of response amplitude. Increases in FFR amplitude with no accompanying decreases in phase locking consistency suggest it is the size of the recruited neural population that is increasing and not a change in the temporal precision of the response

To investigate the effects of immediate task demands and long-term experience on the FFR, this dissertation will investigate the effects and interactions of task demands and expertise on FFR amplitude and consistency. Measures of FFR amplitude and consistency will provide insight into how the previously measured electrophysiological responses from single units behave on aggregate when indexed by the scalp-recorded potential, and how this aggregate activity, when measured at a high frequency, may relate to the late, slow, high-amplitude changes in neural activity measured by BOLD, ERPs or low-frequency oscillatory activity. Gain modulation and entrainment theories predict that selective attention induces an overall increase in FFR amplitude, whereas dimension-selective and dynamic tuning models of auditory processing in the cortex predict a decrease in gross FFR amplitudes with no changes in measures of phase locking consistency. As with other studies of learning and expertise, experiments in this dissertation will use task performance as an index of expertise (Luthra et al., 2024; Gauthier et al., 1998). Specifically, the experiments will compare the responses of high performing participants whose scores are near ceiling with those of lower performing participants. As pre-experiment surveys indicate that task performance is significantly positively correlated with years of musical experience, task performance appears to be a reliable proxy for expertise. Across this dissertation, years of musical experience is measured as most number of years spent playing a single instrument. The main effects of expertise will reveal the influence of training and long-term experience on auditory encoding. Both gain modulation theories and entrainment theories, and dimension-selective and dynamic tuning accounts make similar predictions about the effects of expertise and selective attention. Former theories predict an increase in FFR amplitude and consistency with expertise, whereas latter theories predict a decrease in FFR amplitude but an increase or no change in consistency with expertise. Equally importantly, interaction effects between task demands and task performance may provide a clue as to why previous research on the effects of selective attention on the FFR has been so mixed. Lastly, the main effects of either variable on the FFR with no interaction effects may suggest that these variables influence neural gain or selectivity through independent long-or short-term adaptive mechanisms, and challenge the assumption that greater FFR amplitudes necessarily indicates enhanced auditory encoding.

The second chapter of this dissertation reports on an EEG-FFR study using a paradigm derived from past FFR studies to investigate the potential confounds underlying the historically mixed results in studies on the effects of selective attention on the FFR. Mixed results have been reported across FFR studies using intermodal, dichotic, and passive versus active listening tasks to manipulate selective auditory attention (Galbraith et al., 2003; Hairston et al., 2013; Lehmann & Schönwiesner, 2014; Galbraith & Kane, 1993; Galbraith & Doan, 1995; Galbraith & Arroyo, 1993; Galbraith et al., 1998; Lehmann & Schönwiesner, 2014; Hoorman et al., 2000; Varghese et al., 2015). The relative merits of each manipulation are assessed in the chapter and used to justify the experiment design used in this study. The primary goal of this study was to investigate task performance as a potential confound by comparing interaction effects between task demands and task performance on the FFR. In addition to task performance, this study also investigates stimulus duration as a potential confound in FFR studies and demonstrates both theoretically and experimentally why the stimulus durations used in some past FFR studies have been insufficient to elicit and measure a stable and isolatable FFR response. While this study finds no effect of task demands on the FFR, it finds a strong main effect of task performance on FFR amplitudes, leading to the main conceit and question of the remainder of this dissertation.

The third chapter presents the second EEG-FFR study in this dissertation. This study used a variant of the oddball task or target monitoring tasks commonly used in FFR and auditory selective attention experiments due to its utility in manipulating different dimensions of attention, comparing the effects of task performance and task demands, and monitoring the effects of top-down attention on the FFR by enabling the comparison of FFR signals to well-documented ERPs. In this study, participants listened to sequences of three pure tones played in randomized orders with equal odds. While most auditory oddball tasks use rare complex sounds or white noise bursts as target stimuli, in each trial of this study, one of the three distractors was randomly selected to be the target tone instead to enable the analysis of phase-locked responses to both standards and targets. The additional advantage of this task is that the ERP components associated with different attentional processes involved with the oddball task, such as the N1, P2, and P3b, are well documented (Sutton et al., 1965; Polich, 2007; Luck, 2011; Näätanen et al., 1992). This experiment design allows the effects of task performance on an active listening task to be directly compared to measures of both early phase locking and late, large-amplitude cortical responses to adjudicate between gain modulation and dynamic tuning models of auditory attention.

The fourth chapter addresses the third and most important EEG-FFR study of this dissertation, which compared FFRs in high performers to the same stimuli under passive and active listening conditions. As previous studies of expertise on the FFR have looked only at responses under passive listening conditions, the current experiment enabled the testing of conflicting predictions made by gain modulation and dynamic tuning models on the effects of auditory attention on high-frequency, phase-locked neural responses. This experiment replicated the reduced FFR amplitudes in high performers with the extended musical experience observed in the previous experiment, suggesting that expertise sharpens the neural responses of high performers regardless of context. Importantly, this experiment also found a decrease in FFR amplitude in the active relative to the passive listening task. These results suggest attention engages an increase in selectivity and the dampening of overall neural activity, while passive listening engages a broad, non-selective neural response. Additionally, the active task requires participants to monitor only a specific frequency band for targets and match incoming stimuli to a mental template of target tones, potentially activating a more selective response. Additional analyses run by splitting participants up by performance trajectory enable participants to be grouped based on effort and task performance independently. The effects that are separated by this grouping suggest that effort and expertise are independently associated with different neural mechanisms for enhancing early auditory encoding.

The fifth chapter will address the final experiment presented in this dissertation. This study used the same oddball variant paradigm as the previous two experiments with an additional condition added to manipulate the predictability and therefore the difficulty of the task. The predictability manipulation was added as dimension-selective and dynamic tuning models of auditory attention implicate the ability to anticipate the sensory consequences of upcoming stimuli and reduce their bottom-up encoding as a component of expertise (Holt et al., 2018; Idemaru & Holt, 2020; Fritz et al., 2007). Additionally, this study tests whether the diffuse passive listening mechanisms engaged in the previous experiment are also engaged in trivial active listening tasks. This experiment finds no effects of predictability on FFR amplitudes. However, the decrease in FFR amplitude associated with high task performance was replicated again, suggesting that expertise sharpens subcortical pitch representation regardless of context, but there were also no interaction effects between task performance and predictability. Again, perhaps modulations of selective attention induced by task demands created changes in early auditory processing that were too fine-grained to easily detect. Moreover, the performance effects observed in this and previous studies reveal the importance of distinguishing encoding efficiency and expertise from effortful high performance. Perhaps the effects of task demands that can be observed can be reduced to a matter of overall effort and conscious attention or task difficulty (Fritz et al., 2007). Alternative explanations of this data and necessary follow-up experiments are discussed further within this chapter.

CHAPTER 2

EXPERIMENT 1: EXPLORING TOP-DOWN EFFECTS ON THE FFR

2.1 Introduction

In order to build a more comprehensive understanding of the cognitive role of the frequency-following response (FFR) in audition and auditory attention, the first study in this dissertation was designed as a conservative replication of the top-down attentional effects reported in prior literature using a well-established paradigm (e.g., Galbraith & Kane, 1993; Galbraith et al., 2003; Hartmann & Weisz, 2019). Past findings on attentional modulation of the FFR have been mixed and often weak, raising questions about the robustness of these effects. To address this, the present experiment controlled for several common confounds in prior work and introduced additional analyses to test whether attentional effects emerge more strongly under specific conditions. Specifically, the study examined whether stimulus duration influences the reliability of FFR power estimates. It is predicted that longer stimulus durations are required for an accurate measure of the magnitude of the stable state component of the FFR. In addition, attention effects were evaluated in relation to task performance, testing whether top-down modulation of the FFR is more apparent in high performers for whom attentional manipulations may be more effective. Comparing the effects of attention and expertise helps to probe the role of the FFR in audition and auditory attention. Competing theoretical accounts make divergent predictions about the role of the FFR in the neurobiological instantiation of auditory attention. Gain modulation models, informed by late, low-frequency neural responses measured through BOLD and ERP, posit that attention and expertise enhance auditory encoding by increasing neural gain (Hillyard et al., 1998; Zatorre et al., 1999, 2012). In contrast, dynamic tuning models, supported by intracranial recordings of firing rates in primary auditory cortex, argue that attention and expertise sharpen neural selectivity by suppressing irrelevant activity while maintaining precise phase-locked responses (Fritz et al., 2003; Elhilali et al., 2007). Thus, the present study aimed to test whether attention and expertise recruit gain modulation–like amplification or dynamic tuning–like sharpening in early auditory encoding, developing the framework expanded on across subsequent chapters.

A variety of experimental paradigms have been deployed to probe top-down attention effects in the FFR with mixed effects across them all. Dichotic listening tasks have yielded inconsistent evidence for attentional modulation of the FFR. Some studies have reported enhanced FFRs to the attended stream (Lehmann & Schönwiesner, 2014; Galbraith & Arroyo, 1993), while others have failed to replicate these effects (Varghese et al., 2015) or found them to emerge only under specific signal-to-noise conditions (Price & Bidelman, 2021). Intermodal tasks, which direct attention between auditory and visual modalities, have also shown variable effects. Some early studies (e.g., Galbraith et al., 1998) demonstrated enhanced FFRs to attended auditory stimuli when attention was directed away from competing visual stimuli. Yet, others have reported null effects, particularly in paradigms where auditory stimuli were task-irrelevant or perceptually simple (Hoormann et al., 2000; Coffey et al., 2017). Studies comparing FFRs under passive listening with tasks requiring active discrimination or detection of target tones have reported modest increases in FFR amplitude or consistency (Hartmann & Weisz, 2019; Varghese et al., 2015). These observed effects are often small, and it remains debated whether they reflect true attentional enhancement or confounding factors such as arousal or motor preparation (Bidelman & Momtaz, 2021).

The mixed findings across studies suggest that attention-related effects on the FFR are likely subtle and difficult to detect. One factor that may contribute to this variability is the effectiveness of the attention manipulations, as reflected in task performance. Research in other domains has emphasized the importance of analyzing behavioral performance to assess whether stronger attention effects emerge in individuals who successfully engage selective attention to monitor specific stimulus features (Engelmann et al., 2009; Goh et al., 2014). The FFR studies reviewed may not have considered performance as a secondary task metric due to the small sample sizes used in the majority of such studies. Twenty studies, including those reviewed previously, had an average sample size of 15 participants (SD = 5). On the other side, existing studies that have looked at the effects of expertise on the FFR compare scores on secondary inventories of musical ability or auditory acuity with measures of FFR amplitude and consistency recorded separately, under passive listening conditions (e.g., Parbery-Clark et al., 2009; Strait et al., 2009; 2013, Wong et al., 2007; Musacchia et al., 2008; Krishnan et al., 2005; Skoe et al., 2017; Omote et al., 2017; Russo et al., 2005; Song et al., 2008).

In addition to comparing attention and task performance effects, this experiment addresses the gap between different models of the neural mechanism of auditory attention by comparing measures of FFR amplitude and consistency. Amplitude is typically measured using the magnitude spectrum or power spectrum converted to decibels, and consistency is typically measured by PLV or inter-trial coherence (ITC). Although the two types of measures generally are conflated, they may vary independently as the magnitude spectrum may reflect both neural recruitment and the consistency of phase locking, while PLV indexes the consistency of phase locking alone (Lachaux et al., 1999; Zhu et al., 2013). Higher amplitudes at a particular frequency in the magnitude spectrum indicate strong neural activation but do not necessarily imply consistent phase locking. Gain modulation models, informed mostly by slow changes in BOLD activity generated by large cortical regions, and models such as biased competition and dynamic tuning, which are primarily shaped by intracranial recordings in animals, make different predictions about the effects of expertise and top-down attention on the two measures. Decreases in magnitude without corresponding decreases in PLV suggest that high task performers recruit smaller populations of neurons in the phase locking of stimuli without reducing the consistency of this phase locking. PLV was calculated in these experiments according to Zhu and colleagues (2013) by extracting the phase angles at the frequency of interest from the complex Fourier transform of each trial, then representing each phase as a unit vector on the complex plane. The PLV is obtained by averaging these vectors across trials, with the vector length reflecting the consistency of phase alignment.

Alongside addressing questions of auditory attention, this study was also intended as a conservative replication of some of the top-down attention effects reported in previous experiments to address some of their methodological confounds. One possible confound commonly found in FFR studies of top-down attention that this experiment will address is stimulus duration. A majority of early studies on the FFR that reported no attention effects used incredibly short stimuli. For example, Galbraith and Kane (1993) used at 17.5 msec long 230 Hz tone, Galbraith and Arroyo (1993) used 25 msec long 400 and 200 Hz tones, Hoorman and Falkenstein (1994) used 30 msec long 340 Hz tones, Hoorman and Falkenstein (2000) used 30 msec long 320 and 370 Hz tones, Hoorman and Falkenstein (2004) used 320 Hz 56 msec cue tones. In comparison, studies that have found an effect of intermodal, dichotic, or active attention on the FFR have used comparatively longer stimulus durations. For example, Galbraith and colleagues (1998) used 320 msec long vowels with 170 and 230 Hz F0s, Galbraith and colleagues (2003) used 86 msec long 292 Hz tones, Hairston and colleagues (2013) used 110 msec long 220 Hz probe tones, and Lehmann & Schönwiesner (2014) used 210 msec long vowels with 170 and 225 Hz F0s, Price and Bidelman (2021) used 100 msec synthesized vowels with a 150 Hz F0.

The frequency resolution of the frequency spectrum of a Fourier transform is given by df = Fs/N, where df is the frequency resolution, Fs is the sampling frequency, and N is the number of samples in the signal. Frequency resolution refers to the bin-width of the resulting spectrum, with smaller values indicating a finer resolution. As signal length is inversely proportional to frequency resolution, longer signals can improve resolution. On the other hand, due to the uncertainty principle, while greater sampling rates increase the Nyquist frequency and enable the analysis of higher frequencies, for a fixed number of samples, a larger sample rate will lead to poorer frequency resolution. Another variable to consider is the frequency of the signal of interest. The frequency determines the number of cycles present in a given signal length, with more cycles yielding a better signal-to-noise ratio, and fewer cycles leading to a more uncertain or smeared power estimate. Lower frequency stimuli will therefore require longer signal durations. The review of previous FFR studies finds that a majority of experiments with null results included fewer than 10 cycles of the frequency of interest in their stimulus signals.

To investigate the effect of stimulus length on temporal resolution, this study employs pure tone stimuli of varying lengths. For each trial in the auditory condition, participants listened to sequences of tones consisting of either the 90 Hz or the 220 Hz pure tone standards. The 90 Hz tones standards were either 120 or 220 msec long, while the 220 Hz tones were either 50, 90, or 120 msec long. One standard tone was selected for every trial to be played repeatedly in a sequence, and standards were occasionally randomly replaced with rare target tones that were either slightly higher or lower in pitch than standards. Depending on randomized instructions, participants were asked to either count all the high targets or all the low targets. The tone lengths were selected such that each standard tone had one length consisting of approximately 10 cycles of the signal and one length consisting of approximately 20 cycles of the signal. Interstimulus intervals were maintained at a consistent 100 msec. These stimulus frequencies were selected based on their correspondence with response peaks reported by Tichko and Skoe (2017), where peaks in lower frequency ranges in the FFR power spectra reflect greater cortical activity and peaks in higher frequency ranges reflect greater contributions from subcortical areas.

As with the other studies in this dissertation, consistently high task performance, which is significantly correlated with musical experience, is used here as a measure of expertise. Years of musical experience was measured as the self-reported longest duration spent playing a single instrument. Gain modulation models predict that attention and expertise should both increase neural activity in response to attended stimuli and hence increase FFR amplitude and consistency. In contrast, dynamic tuning models predict an increase in response selectivity, explaining the changes observed in fMRI data as relative increases in activity driven by decreases in non-feature selective brain regions. Dynamic tuning models therefore predict an increase in consistency accompanied by decreases in amplitude from both expertise and attention. Interaction effects between task performance and task demands are predicted, with the greater differences in FFR between auditory and visual tasks in higher auditory task performers. Given that task performance effects are more consistently observed than attention effects, there may be a main effect of performance without any interactions with task demands. This would suggest that expertise sharpens the neural responses regardless of context. Lastly, it is predicted that FFR amplitudes will be greater with longer stimulus durations, demonstrating the necessity of using longer stimulus signals in FFR studies.

2.2 Materials and Methods

Twenty participants were recruited from the University of Chicago community (14 female-identifying, 5 male-identifying, and 1 non-binary) with ages ranging from 20 to 27 years old. All participants had normal speech and hearing, as verified with the Welch Allyn Hearing Screener (Welch Allyn, User Manual, 2013). Participants signed an informed consent form reviewed by the Institutional Review Board of the University of Chicago and were compensated with cash at \$20/hr or with course credits. Participants completed a survey on their language and music experience prior to the start of the experiment.

Stimuli and task

Participants were presented with two simultaneous streams of visual and auditory stimuli on a PC running Ubuntu Studio 20.04 using the Python package PsychoPy (Peirce et al., 2019). The auditory task consisted of sequences of pure sine-wave tones of either 90 or 200 Hz. Tone onsets and offsets were ramped with a 5 ms Hanning window. The 90 Hz tones were either 120 or 220 ms, while the 220 Hz tones were either 50, 90, or 120 ms, resulting in 5 possible standards. The interstimulus interval between all tones was a constant 100 ms. One standard tone was randomly selected for each trial, for example, in some trials, all the standard tones were 90 Hz tones, 220 ms in length, while in other trials, all the standards were 220 Hz tones, 90 ms in length. Participants were asked to listen to the tones and count every time a standard was replaced by a slightly higher or lower target tone, which appeared 5% of the time. Target tones for the 90 Hz condition were 5 to 15 Hz higher or lower than the 90 Hz standards, while target tones for the 220 Hz condition were 10 to 30 Hz higher or lower than the 220 Hz standards. Sequences consisted of 105 tones each, with 30 trials per block. At the end of each trial, participants were asked to report their counts and received a point for each correct answer. To incentivize performance, participants were paid an additional $10 if they scored high enough, although all participants received this bonus. During the visual task, participants were instructed to complete a task involving rapid rotations of the fixation cross. In some trials, participants were asked to count every time the fixation cross rotated in a clockwise direction, while in other trials, participants were asked to count every counterclockwise rotation. Rotations were limited to 5 to 30 degrees from the center on either side. Visual targets appeared during the interstimulus interval of tones and remained visible for 30 ms. Visual and auditory stimulus streams were presented simultaneously, and auditory and visual tasks were presented in separate blocks. There were four total experiment blocks, and participants received two blocks of each task in an alternating order, such that all odd-numbered participants received the visual blocks first and third, and even-numbered participants received the auditory blocks second and fourth. Participants received approximately 1000 sweeps of each standard tone under each condition. The task lasted approximately 80 minutes.

Electrophysiological data acquisition and preprocessing

Prior to EEG cap placement, participants were briefed on the experimental procedure and what to expect during the EEG recording. Head circumference was measured to determine the appropriate size of the 64-channel actiCAP slim net (actiCAP, Brain Products GmbH, Germany). Participants were seated in a well-lit, sound-attenuated, and electromagnetically shielded room and instructed to attend to the auditory stimuli while minimizing eye blinks and other movements. Following the experiment, electrode positions were recorded using the CapTrak system (Brain Products GmbH, Germany). EEG signals were acquired using an actiCHamp Plus amplifier (Brain Products GmbH, Germany) at a sampling rate of 5 kHz. Two electrodes, positioned at sites corresponding to AF7 and AF8 in the actiCAP layout, were placed near the outer canthi of the eyes to monitor ocular activity and serve as electrooculogram (EOG) channels. EEG preprocessing was conducted using custom Python scripts and the MNE-Python library (v1.5.0; Gramfort et al., 2013). Eye electrodes (AF7 and AF8) were re-referenced to Fp1 and Fp2, respectively, creating bipolar EOG channels (i.e., referenced from below the eye to the contralateral electrode above the eye). The automated PREP pipeline was applied to remove 60 Hz line noise using a notch filter, identify and exclude noisy channels based on amplitude thresholds, and re-reference the EEG signals to the common average. Before both experiments, a timing calibration test was performed, and the mean event timing lag was measured to be 27.2 ms with a jitter of 0.73 ms.

FFR analysis

Since tasks in this experiment used only stimuli of a single polarity, the common average reference applied during PREP preprocessing was maintained. Preprocessed data from all channels except Cz were dropped for FFR analysis. Using a common average reference mitigates the likelihood of biasing responses towards peripheral sources, including the cochlear microphonic and auditory nerve, and overestimating the absolute amplitude of the signal (Bidelman et al., 2015; Coffey et al., 2019). The data were epoched from -100 ms to the stimulus duration relative to stimulus onset. The epoched data were then time shifted by -27.2 ms based on the lag in stimulus presentation calculated as the difference between the timestamp of each event tag and the onset of the stimulus signal as recorded by the StimTrak (fig. 2). The epochs were then bandpassed from 50 to 300 Hz using an FIR filter with a lower and higher transition bandwidth of 12.5 Hz and 75 Hz and a filter length of 6.6 times the lower transition bandwidth. To compute the power at each stimulus frequency for each participant, trials for each tone and target condition were first averaged in the time domain. Evoked potentials were then cropped to the stimulus length with an additional 20 ms of buffer time. A Hanning window of the same length as the baseline was applied over the baseline period. The windowed signal was then zero-padded to match the length of the evoked signal. Finally, the ratio of power between the two signals was computed in decibels (dB). This analysis was repeated for the second harmonic of the 90 Hz tone at 180 Hz. FFRs at the second harmonic of the 220 Hz tone were not analyzed because the zero padding of the baseline period increased the noise floor, especially for higher frequencies.

PLV analysis

The consistency of phase locking in the neural response was calculated using the PLV (Lachaux et al., 1999; Zhu et al., 2013). The epoched FFR signals were cropped to keep only the segment of each epoch following stimulus onset. The PLV for each participant and trial was computed using bootstrapping; 400 trials were randomly drawn from each condition, and the fast Fourier transform was computed for each epoch. Next, the phase information at each stimulus frequency was extracted by returning only the angle of the complex-valued output of the FFT. Finally, the PLV for the draw was computed by averaging the complex exponential of the phases from all 400 randomly drawn trials. This process was repeated 1000 times to compute the average PLV for each stimulus for each participant. A null distribution generated by phase-scrambling provides a robust baseline for statistical testing. Then, the null distribution was computed per Zhu and colleagues (2013) by calculating the PLVs over 1000 randomly drawn phases and repeating this process 1000 times. The significance of the observed PLV values was assessed by conducting pairwise tests on the observed PLVs and the simulated null, as well as by comparing the distributions of observed PLVs for high- and low-performance participants.

2.3 Results

Behavioral results

Participants performed well at both tasks with a few at ceiling. Participants had slightly higher accuracies for the auditory task (Mdn = 82%, IQR = 26%) than the visual task (Mdn = 86%, IQR = 16%) (W = 53, $p < 0.1$) (Figure 1). Although there were a few outliers for both the visual and the auditory task, none of the participants had anomalously low scores for both tasks so their data was not excluded. Participants were separated into high and low auditory task performance groups based on a median split for later analyses. Musical experience was measured as the number of years spent playing a single instrument. An Ordinary Least Squares regression was run between years of musical experience and performance accuracy in the auditory. The model was weakly statistically significant F(1, 20) = 3.03, $p < 0.1$, $R^2 = 13.8\%$ with years of musical experience significantly predicting

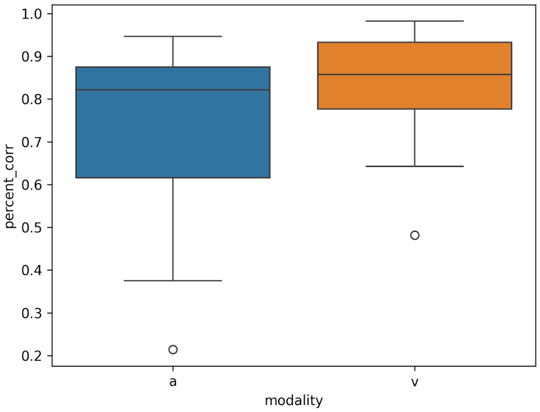


Figure 2.1: Participants performed slightly better in the visual than in the auditory task. Performances were high across both tasks, but very few participants were at ceiling. There were a few participants with anomalously low scores for either task, but no participant had anomalous scores for both tasks, so their data were not excluded.

scores b = 1.3%, $p < 0.1$.

FFR results

A linear mixed-effects model with random intercepts was fitted using lme4 (Bates et al., 2015) on R to assess the effects of task modality (active vs. passive), stimulus duration (0.05, 0.09, 0.12 s), and auditory task accuracy (high vs. low performers) on FFR amplitudes to the 220 Hz tone. Separate models were used for each tone because stimulus duration and stimulus frequency were not fully crossed in this study. Subject was included as a random intercept to account for repeated measures. Post-hoc comparisons were conducted using the emmeans package on R (Lenth, 2025) (Figures 2.2 and 2.3).

The model revealed a significant main effect of stimulus duration. Relative to the shortest duration (0.05 s), FFR amplitudes were significantly larger for 0.09 s (b = 7.22, SE = 1.32, t(90) = 5.48, $p < .0001$) and 0.12 s stimuli (b = 10.63, SE = 1.32, t(90) = 8.07, $p < .0001$). Estimated marginal means showed a monotonic increase in amplitude with longer durations: M = 8.05 μV (95% CI [5.27, 10.8]) at 0.05 s, M = 15.31 μV (95% CI [12.53, 18.1]) at 0.09 s, and M = 17.59 μV (95% CI [14.81, 20.4]) at 0.12 s. Post hoc Tukey-corrected comparisons confirmed significant differences between all duration pairs: 0.09 s > 0.05 s (estimate = 7.26, SE = 0.66, t(90) = 11.02, $p < .0001$), 0.12 s > 0.05 s (estimate = 9.54, SE = 0.66, t(90) = 14.47, $p < .0001$), and 0.12 s > 0.09 s (estimate = 2.28, SE = 0.66, t(90) = 3.46, p = .0024).

By contrast, there were no significant main effects of task modality (p = .92) or auditory accuracy group (p = .89). The interaction between duration and accuracy approached significance at 0.12 s (b = –3.45, SE = 1.86, t(90) = –1.85, p = .067), with high performers showing numerically greater responses than low performers (M = 19.46 vs. 15.72 μV), with this different reaching slight statistical significance (estimate = 3.73, SE = 2.17, t(23.3) = 1.72, p = .098). No other two- or three-way interactions were significant (all ps > .36). Together, these findings indicate that FFR amplitudes to the 220 Hz tone increased robustly with stimulus duration, and that task performance slightly increases FFR amplitudes, with this effect reaching significance only for longer stimulus durations.

A separate mixed-effects model with random intercepts was fitted for the FFRs to the 90 Hz tone. This model included fixed effects for task modality (active vs. passive), stimulus duration (0.12, 0.22 s), auditory task accuracy (high vs. low performers), and their interactions, with subject modeled as a random intercept. The analysis revealed a significant main effect of stimulus duration, b = 4.77, SE = 1.52, t(54) = 3.13, p = .0028. Estimated marginal means indicated that FFR amplitudes were higher for the 0.22 s stimuli (M = 16.0 μV, 95% CI [13.7, 18.2]) than for the 0.12 s stimuli (M = 10.8 μV, 95% CI [8.56, 13.1]). Tukey-corrected post hoc contrasts confirmed this difference (estimate = 5.14, SE = 0.76, t(54) = 6.74, $p < .0001$). There were no significant main effects of task modality (p = .79) or auditory accuracy group (p = .70), and no significant interactions among these factors (all ps > .26). Thus, FFR amplitudes to the 90 Hz tone were primarily modulated by stimulus duration, with longer tones eliciting stronger responses, while task engagement and performance level did not significantly influence responses under the present conditions.

Finally, a separate mixed-effects model was fitted to examine FFR amplitudes at the second harmonic (180 Hz) of the 90 Hz fundamental frequency. This model included fixed effects for task modality, stimulus duration (0.12, 0.22 s), auditory accuracy group, and all interactions, with subject as a random intercept. The analysis revealed a significant main effect of stimulus duration, b = 4.76, SE = 1.73, t(54) = 2.76, p = .0079, indicating higher FFR amplitudes for the longer 0.22 s stimulus (M = 11.91 μV, 95% CI [10.16, 13.7]) compared to the shorter 0.12 s stimulus (M = 8.57 μV, 95% CI [6.81, 10.3]). However, there were no significant main effects or interactions involving task modality (all ps $>$ .26) or auditory task accuracy (all ps $>$ .13), suggesting that neither active engagement nor performance level modulated neural encoding at the second harmonic. These findings contrast with prior studies reporting attentional effects at harmonic frequencies and suggest that the FFR at the second harmonic may be less sensitive to top-down modulation under the current paradigm.

PLV results

PLVs were computed across each condition to compare measures of FFR amplitude with measures of phase locking consistency (Figure 2.5). A bootstrapped null distribution of randomized PLV values was used to determine whether the observed PLV values are statistically significant. The observed values are compared to the null distribution through percentile-based thresholding. Using a significance level of 0.05, the computed 95th percentile of the null distribution is 0.056. PLV values for both high (Mdn = 0.096, IQR = 0.041) and low performers (Mdn = 0.071, IQR = 0.049) to the 90 Hz stimuli are significantly greater than the threshold, indicating that the observed PLV values did not occur due to chance ($p < 0.05$). Likewise, PLV values for both high (Mdn = 0.181, IQR = 0.133) and low performers (Mdn = 0.174, IQR = 0.122) in response to the 220 Hz stimulus are significantly greater than null ($p < 0.05$). Observed PLV values were within the range of observed PLV values in other FFR studies (e.g., Omote et al., 2017; Doelling & Poeppel, 2015; Holmes et al.,

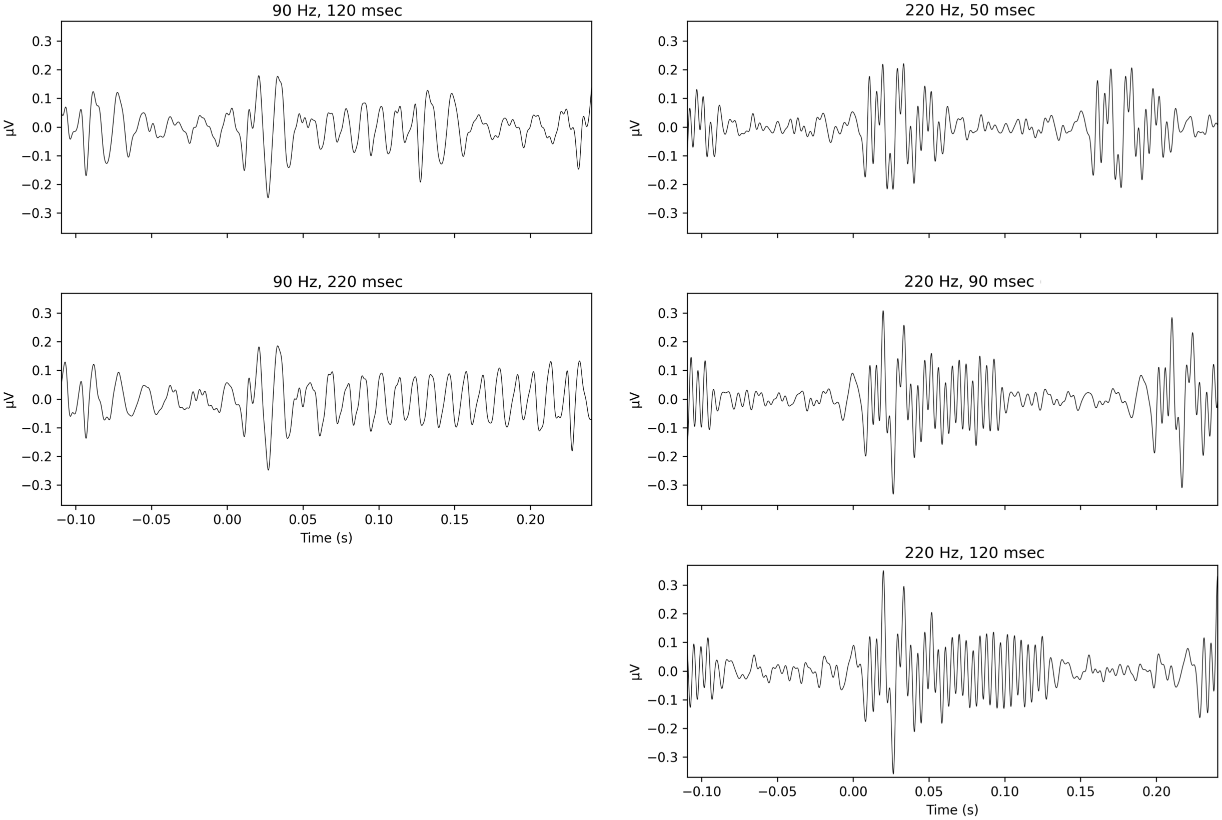


Figure 2.2: Grand-average time domain FFRs to each standard tone across both conditions. Stimulus onset occurs at zero seconds. Both the transient brainstem response and the stable-state phase locked response can be seen. The transient portions recorded here last up to approximately 50 ms after stimulus onset, occluding the stable state response. Responses can also be seen to extend slightly past stimulus offset.

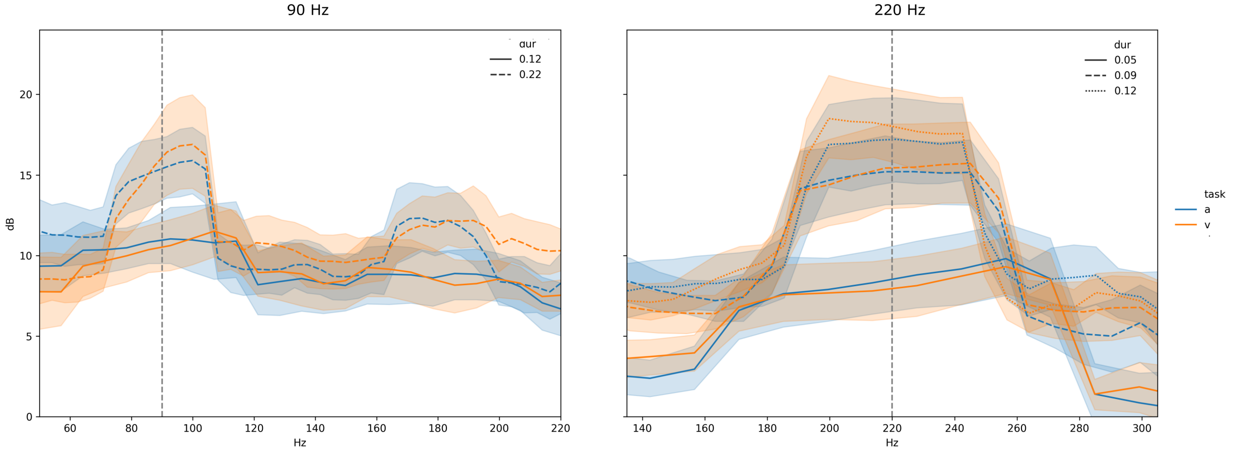


Figure 2.3: Amplitude spectrum for each standard tone. There is no effect of task modality on FFR amplitudes. However, greater FFR power and sharper peaks are observed for FFRs to longer tones. The frequency resolution is a direct function of the sampling rate and signal length. Therefore, the 90 Hz, 120 ms standard and the 220 Hz, 120 ms standard have the same frequency resolution and peak width. Shaded regions represent the 95% CI.

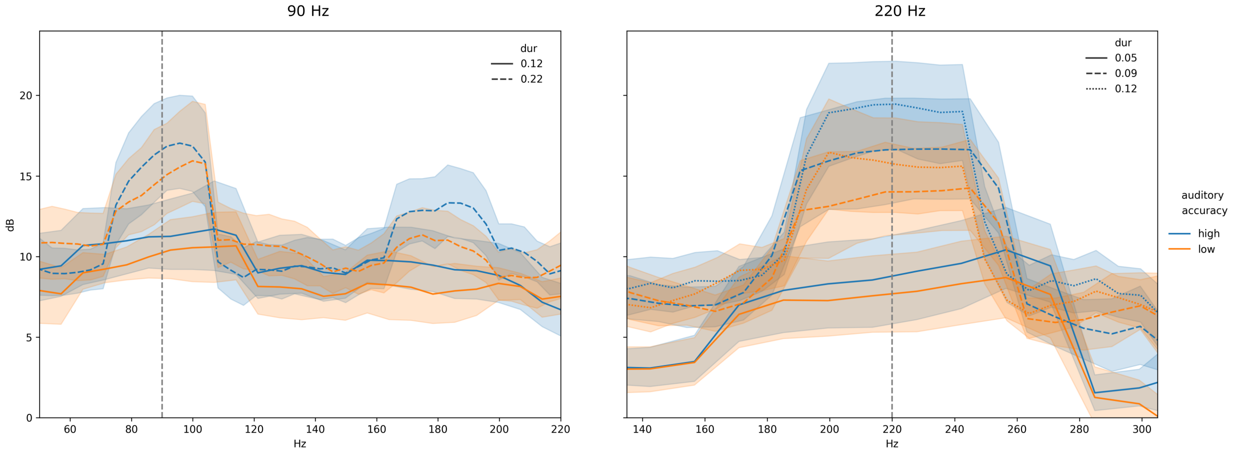


Figure 2.4: FFRs were grouped by high and low performers on the auditory and visual tasks. FFR amplitudes are significantly higher for high task performers for the 220 Hz tone and at the second harmonic of the 90 Hz tone. Lines represent the mean power at each frequency, and shaded regions represent the 95% CI.

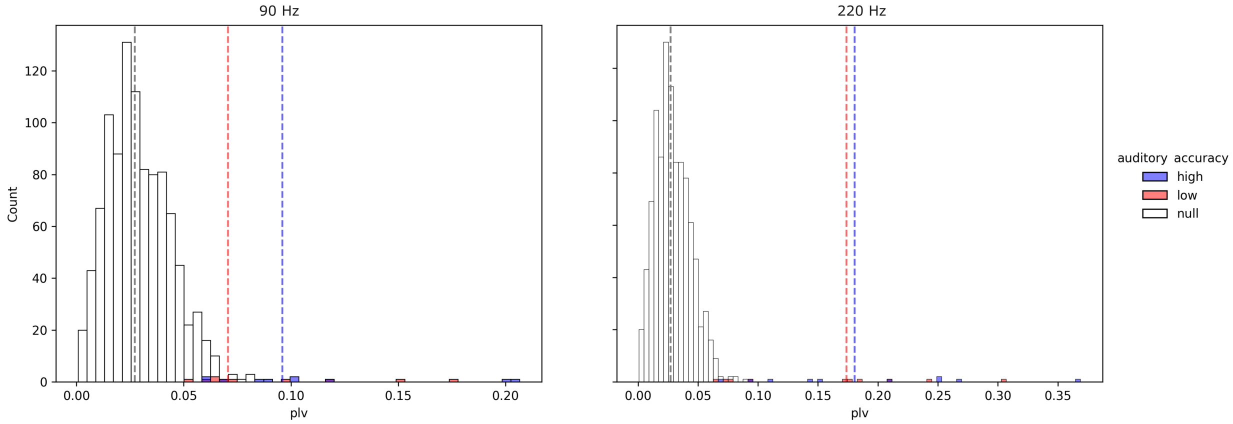


Figure 2.5: Observed PLV values were compared to a simulated null distribution. PLVs for both high and low task performers are greater than null across both tone frequencies. PLVs between groups were not significant. Dotted lines indicate the median value of each distribution. To enable a clear depiction of both the null and observed distributions only 500 values from the null distribution were used in these figures.

2018). Wilcoxon Signed-rank tests between each performance group for each tone revealed no significant differences ($p > 0.1$).

2.4 Discussion

This study investigated the extent to which task demands, task performance, and stimulus duration influence auditory frequency-following responses (FFRs), using both amplitude and phase-locking value (PLV) as dependent measures. While past work has provided mixed evidence for top-down modulation of the FFR, our results suggest that task performance, but not task demands, is significantly associated with changes in FFR amplitudes. Specifically, individuals who performed better on the auditory task exhibited stronger FFR amplitudes with no change in PLVs, indicating a more robust neural tracking of the periodic stimuli. In contrast, no main effects of attentional task condition were observed, and interaction effects between task demands and task performance were absent, suggesting limited effects of selective attention on FFR properties. Additionally, FFR amplitudes scale with stimulus duration, particularly for longer signals, highlighting the importance of longer stimulus signals in future FFR studies.

Top-down influences on the FFR

The primary finding of this study was a correlation between FFR amplitude and high task performance. High-performing participants exhibited significantly greater FFR amplitudes to both 90 Hz and 220 Hz stimuli, suggesting greater phase locking to the stimulus signals. These results extend earlier findings that link auditory expertise or musical training with increased FFR fidelity (e.g., Parbery-Clark et al., 2009; Wong et al., 2007; Strait et al., 2013) by showing that even short-term task performance may reflect underlying differences in auditory processing efficiency. Importantly, performance was measured within the same session and paradigm, whereas past FFR studies demonstrated increased FFR amplitudes and consistency in experts' recorded electrophysiological responses exclusively under passive listening conditions. In combination, these results suggest that FFR amplitudes and consistency are higher in experts regardless of context.

The finding that participants with higher task accuracy exhibited increased FFR amplitudes and greater phase consistency (PLV) supports gain modulation accounts of auditory attention. According to gain-based models, attention enhances neural processing by increasing the responsiveness of neurons tuned to task-relevant features (Zatorre et al., 2002). This amplification results in larger neural responses to attended stimuli, a phenomenon widely observed in studies using fMRI and MEG, where auditory attention leads to enhanced activity in cortical regions such as Heschl's gyrus and the planum temporale (Pantev et al., 1998; Zatorre et al., 2012). The current results suggest that this attentional gain may extend to earlier stages of auditory processing, as indexed by the FFR, which reflects both subcortical and cortical phase-locked neural activity. Increased FFR amplitude in high performers implies that heightened neural activity is associated with greater attentional engagement and enhanced encoding of the periodic acoustic features. These findings are consistent with gain modulation theories that predict both stronger and more temporally aligned responses under conditions of focused attention and successful task performance.

The present findings do not align with dynamic tuning models of auditory attention, which propose that attentional mechanisms enhance perception not by increasing overall neural activity, but by sharpening neural selectivity and suppressing responses to irrelevant features. According to this framework, expertise or focused attention should lead to more selective neural firing, resulting in reduced overall response amplitudes but enhanced signal fidelity and phase locking consistency (Fritz et al., 2007; Holt et al., 2018). However, in this study, participants with higher auditory task accuracy exhibited increased FFR amplitudes, suggesting a global enhancement in neural responsiveness rather than a narrowing of selectivity. The absence of any amplitude attenuation contradicts the core prediction of dynamic tuning models, indicating that attentional or performance-related improvements in FFR encoding are better explained by gain-based mechanisms than by tuning-driven suppression of non-relevant activity.

Contrary to some previous findings, the present study found no main effects of attentional task condition on FFR amplitude or PLV. This null result echoes other studies that have failed to find attention effects in intermodal paradigms (Hoormann et al., 2000; Coffey et al., 2017), particularly when auditory stimuli are simple tones and not speech-like signals. One interpretation is that task-based attentional modulation of the FFR is subtle and difficult to detect without strong attentional manipulation or complex auditory stimuli. While participants were asked to selectively attend to either an auditory or a visual stream, the intermodal task may not have imposed sufficient attentional load or competition to strongly modulate subcortical or early cortical processing, especially in the context of simple tone stimuli. Additionally, the high overall performance across participants in both tasks may have limited the range of attentional engagement needed to detect meaningful interactions with expertise. Prior work has shown that attention effects on FFRs are often small and highly context-dependent (Coffey et al., 2017; Price & Bidelman, 2021), emerging more reliably in paradigms with high task difficulty, speech stimuli, or more distinct auditory streams. These findings therefore underscore the challenge of using relatively fine-grained manipulations in intermodal paradigms to capture subtle attentional modulations of early auditory responses.

The presence of a strong association between task performance, which was significantly correlated with musical experience, and the absence of any main or interaction effects of task demands, suggests that expertise exerts a broad, context-independent influence on early auditory processing. This pattern implies that individuals with higher auditory task performance may possess higher neural gain or sensitivity regardless of context. Rather than reflecting transient, top-down attentional modulation, the observed enhancements in FFR amplitude and PLV may reflect stable changes in synaptic plasticity shaped by long-term experience. These long-term changes may be carried out by long-term potentiation (LTP), resulting in the increased excitability of neurons to specific auditory features and greater response amplitudes. LTP has been demonstrated in the auditory cortex and subcortical auditory pathways (e.g., Tzounopoulos & Kraus, 2009). These changes may also be supported by enhanced top-down corticofugal feedback in experts. The observed effects of task performance suggest a general strengthening of this feedback loop rather than task-dependent top-down adaptations.

The role of stimulus duration

Stimulus duration emerged as a robust factor influencing FFR amplitude. Longer stimuli consistently produced higher FFR amplitudes, supporting the theoretical prediction that longer signals improve signal-to-noise ratio and frequency resolution. This finding adds to growing evidence that short stimuli may yield weaker or more variable FFRs, and helps explain why earlier studies using very brief tones often reported null results (Galbraith & Kane, 1993;

Hoormann et al., 1994). From a signal-processing standpoint, longer signals contain more cycles of the frequency of interest, which not only enhances the Fourier representation but also enables better averaging and phase consistency. Particularly for lower-frequency stimuli (e.g., 90 Hz), longer durations are crucial to ensure enough cycles for reliable spectral estimation. Thus, FFR researchers should carefully consider stimulus length when designing paradigms, especially when investigating low-frequency periodicities.

Conclusion

This study provides new evidence that changes in FFR amplitudes are associated with increases in task performance, but not with changes in attentional condition. High performers in an auditory task exhibited stronger and more consistent FFRs, indicating that individual differences in attentional control influence early auditory encoding in a context-independent manner. Stimulus duration also emerged as a crucial design variable, with longer stimuli producing more robust FFRs. Together, these findings highlight the importance of accounting for performance metrics and stimulus parameters in FFR research. Future studies should continue to explore how attention and expertise shape the temporal fidelity of auditory responses, using both amplitude and PLV-based measures. Combining EEG-FFR with other modalities such as fMRI or MEG may help disentangle cortical and subcortical contributions. Moreover, longitudinal studies tracking attentional training or auditory learning could test whether FFR enhancements are causally linked to improvements in auditory performance. Ultimately, a more nuanced understanding of how attention and encoding precision interact may help unify competing theories of auditory cognition.

CHAPTER 3

EXPERIMENT 2: NEURAL RESPONSE SELECTIVITY IN EXPERTS UNDER REDUCED TASK DEMANDS

3.1 Introduction

The previous chapter presented a study investigating the effects of task performance and task demands on the FFR. The presence of main effects on FFR amplitudes and consistency, with no other main effects or interactions with task demands provided evidence supporting long-term, training-induced and context-independent changes in gain modulation in auditory processing. The study presented in this chapter takes a slightly different tact to bridging the gaps between models of auditory processing. Based on the mixed effects seen in the literature on attention effects in the FFR and the absence of attention effects in the previous study, this study uses a paradigm that enables the simultaneous recording of ERPs and FFRs. Recording ERPs and FFRs to the same stimuli enables the direct comparison of fast, phase-locked FFR responses with the slow, large-amplitude responses measured by ERPs cited by auditory models of gain modulation (Hillyard et al., 1973; Näätänen et al., 1978). This study also establishes a paradigm and methodology used in later chapters.

Understanding how attention shapes auditory perception requires tools capable of tracking neural encoding at both short and long timescales. Existing theories of auditory attention, such as gain modulation, biased competition, and dynamic tuning, propose different mechanisms by which selective attention enhances the neural processing of behaviorally relevant auditory information. Yet, empirical investigations tend to rely on either slow, spatially broad neuroimaging measures, such as fMRI, or highly focal, time-sensitive measures, like intracranial electrophysiology. The frequency-following response (FFR) offers a unique bridge between these domains. The FFR is a scalp-recorded potential that reflects phase-locked activity to periodic auditory stimuli, with high temporal precision and sensitivity to both subcortical and cortical generators (Skoe & Kraus, 2010; Bidelman, 2015).

A significant limitation in the current literature on auditory attention and FFRs is the narrow focus on either expertise or task demands in isolation. Most FFR studies examining the effects of expertise, such as musical training, have recorded responses during passive listening conditions (e.g., Wong et al., 2007; Musacchia et al., 2008; Parbery-Clark et al., 2009). While these studies suggest that long-term auditory training can enhance FFR amplitude and phase consistency, they provide limited insight into how such expertise interacts with task engagement or attentional focus. Conversely, studies examining task effects on the FFR often use paradigms where participants are instructed to focus on or ignore auditory stimuli (e.g., Galbraith & Arroyo, 1993; Lehmann & Schönwiesner, 2014; Price & Bidelman, 2021), but these effects are typically small or inconsistent, and rarely analyzed in the context of performance variability or long-term training. Thus, the field lacks paradigms that simultaneously manipulate task demands and assess individual differences in attention or expertise.

To address this gap, the present study employs a modified auditory oddball paradigm designed to examine how top-down attention and auditory expertise jointly influence early and late stages of auditory neural encoding. Specifically, this paradigm was designed to enable the concurrent measurement of FFRs and ERPs to the same stimuli. This approach enables the comparison of subcortical fidelity and cortical discrimination processes within the same task and participant group. It also enables testing of how task demands and individual differences in performance interact to modulate the neural representation of auditory events. The task design introduces three equiprobable pure tones within each trial, with one randomly selected as the target for that trial. This variation departs from conventional oddball paradigms, which often use complex stimuli or white noise bursts, by allowing target and non-target stimuli to share acoustic properties, thus supporting balanced averaging for FFR analysis. Notably, the task elicits well-characterized ERP components, N1, P2, and P300, that are associated with early sensory encoding, perceptual categorization, and attentional updating, respectively (Näätänen et al., 1992; Luck, 2012; Polich, 2007). This allows for direct comparisons between simultaneously recorded early phase locked responses and later cortical processing markers within a common attentional framework to the same stimuli. As before, expertise was determined by task performance, which was deemed an appropriate proxy given the correlation between performance and musical experience, as well as the high proportion of high performers who reached the task ceiling.

The oddball paradigm has long been a staple of ERP research on attention. The N1 component reflects early cortical encoding and is typically enhanced by attention (Hillyard et al., 1973; Näätänen et al., 1992). The P2 has been linked to early perceptual learning and may reflect stimulus evaluation or auditory categorization (Crowley & Colrain, 2004; Luck, 2014). The P300, or more specifically, the P3b subcomponent, is widely accepted as a marker of attentional resource allocation and context updating in working memory (Sutton et al., 1965; Polich, 2007), it has also been associated with the maintenance of an attentional trace of relevant stimuli, and with comparisons between incoming stimuli and this attentional trace (Alho et al., 1987; Alho et al., 1992; Woods & Alain, 2001). Using a paradigm that elicits these components in response to simple, periodic stimuli allows for the direct comparisons of ERP modulations with the FFR, which captures neural tracking of the fundamental frequency of the auditory signal at early (potentially subcortical) stages.

By comparing ERP and FFR responses within the same task and stimuli, this study investigates how attention and expertise jointly shape auditory encoding across the neural hierarchy. The simultaneous recording of ERPs and FFRs enables the examination of whether effects of attention and expertise manifest early, late, or at multiple stages of auditory processing. For instance, gain modulation models predict that attention should increase neural activity across both early and late components (Zatorre et al., 2002). Specifically, gain modulation models predict parallel increases in neural signal strength across measures, including both ERP amplitudes (e.g., larger N1, P2, and P3 components) and FFR amplitude and phase consistency (PLV). In contrast, dynamic tuning models posit that attention and experience refine the selectivity of sensory responses, enhancing the signal-to-noise ratio not through overall gain, but by suppressing activity from irrelevant populations while sharpening responses to task-relevant features. Dynamic tuning thus predicts increased PLV without necessarily increasing FFR amplitude, and does not make direct predictions about ERP components, which reflect broader cortical population activity rather than frequency-specific neural tuning. By testing both FFR amplitude and PLV, as well as ERP amplitude and latency, this paradigm allows direct evaluation of these competing accounts.

Another novel contribution of this study is the introduction of a reusable and flexible experimental design that can be adapted to manipulate local stimulus statistics and attentional demands. The three-tone oddball structure permits fine-grained adjustments to tone probabilities, pitch separation, and target identity without altering the core timing or stimulus features. It also enables reliable measurement of behavioral performance, which can be directly correlated with neural metrics. This is particularly important for assessing the role of individual differences in auditory attention, as recent work has suggested that performance-based metrics may more sensitively index attentional engagement than task instructions alone (Goh et al., 2014; Volosin & Horváth, 2020).

This chapter presents an adapted experimental paradigm that offers a powerful tool for probing the neural mechanisms of auditory attention. By combining FFR and ERP measures, and comparing the effects of both task performance and task demands within the same design, the study aims continues to explore the deployment of gain modulation and dynamic tuning as mechanisms of auditory attention. The task will serve as the foundation for subsequent experiments in this dissertation and is intended to provide a generalizable platform for future research in auditory neuroscience.

3.2 Materials and methods

A total of 45 participants were recruited for this study from the University of Chicago community (11 male, 1 non-binary, with ages ranging from 20-50 years old. All participants had normal speech and hearing as verified with a Welch Allyn otoscope. A further 3 participants were excluded from further analysis due to the high number of trials rejected following the PREP preprocessing pipeline (Bigdely-Shamlo et al., 2015), which utilized a peak-to-peak rejection threshold of 35 mV on the channel used to compute FFRs. The resulting dataset for the final analysis includes a total of 41 participants. Prior to their participation, participants signed an informed consent form approved by the Institutional Review Board at the University of Chicago. Participants were compensated with either cash at $20/hr or with course credits.

Stimuli and task

Stimuli were presented on a PC running Ubuntu Studio 20.04 using the Python package PsychoPy (Peirce et al., 2019). Stimuli were 3 pure sine-wave tones at 130, 200, 280 Hz played in a random order in sequences 30, 36, or 42 tones in length. Tones onsets and offsets were smoothly ramped with a 5 ms Hanning window. The tones were 300 ms in duration with a jittered interstimulus interval from 200-300 ms. Stimuli were presented at 65–70 dB SPL over insert earphones (Etymotic ER3C). At the start of each trial, one of the three tones was randomly selected to be the target tone and participants were allowed to listen to the target tone as many times as they wished. Participants were then asked to silently count the number of times the target tone played during the trial and report their answer at the end of the sequence. The tones were pseudo-randomized such that the first tone was never a target tone, and no tone would play more than three consecutive times. To incentivize task compliance, participants were given one ‘point’ if they answered correctly, or if their guess came within 2 tones of the actual target count. Participants received feedback at the end of each trial in the form of the number of actual targets and their total points. They were informed that they could earn up to $10 bonus in cash at the end of the experiment, depending on task performance, although all participants eventually performed well enough to be awarded the $10. Each block had 18 sequences and there were 4 total blocks with each lasting approximately 8-10 minutes. The entire task lasted approximately 1 hour. This resulted in approximately 850-1000 sweeps per tone.

Participants were asked to silently count the number of targets and report their answer at the end of each trial. Behavioral performance for each trial was therefore calculated as 1- the ratio between the error and the actual number of targets. The overall accuracy for each participant was calculated as the mean accuracy across all trials. The median accuracy was calculated and participants with accuracies below the median were labeled as low performers and participants with accuracies above the median were labeled as high performers.

Electrophysiological data acquisition and preprocessing

All EEG recording and data processing procedures are the same as those used for Experiment 1. Prior to EEG cap placement, participants were told what to expect from the task and EEG procedure. The circumference of each participant's head was measured to fit the actiCAP slim EEG 64-channel net (actiCAP, Brain Products GmbH, Germany). Participants were seated in a bright, sound-attenuated, and electromagnetically shielded room and asked to attend to the tones while minimizing eye blinks and other movement artifacts. After the experiment, the exact electrode positions were recorded using CapTrak (Brain Products GmbH, Germany). EEG data was collected with a sampling rate of 5 kHz using an actiCHamp Plus amplifier (Brain Products GmbH, Germany). Two electrodes (originally AF7 and AF8 in the actiCAP layout) were affixed under the outer canthi of the left and right eyes to monitor eye movements (i.e., to be used as EOG channels).

The EEG data were preprocessed using custom Python scripts and the MNE-Python library version 1.5.0 (Gramfort et al., 2013). The eye electrodes (AF7 and AF8) were re-referenced to Fp1 and Fp2, respectively (i.e., below the eye referenced to above the contralateral eye), to create bipolar EOG channels. The automated PREP pipeline was used to notch filter the data at 60 Hz, identify and exclude bad channels using threshold-based artifact rejection, and re-reference the EEG data to the average of all electrodes. Detailed participant-level preprocessing reports containing details such as trial counts, removed Independent Components, interpolated electrodes, and data visualizations for quality check are available with our dataset on OpenNeuro. In both experiments, a timing test was run before the experiment and the measured lag in event times was 27.2 ms with a jitter of 0.73 ms.

FFR analysis

The FFR and PLV analyses used in this experiment are the same as those used in Experiment 1. The common average reference (CAR) applied during PREP preprocessing was preserved to reduce the risk of biasing the recorded response toward peripheral generators, such as the cochlear microphonic or auditory nerve, and avoids overestimating the absolute amplitude of the FFR signal (Bidelman et al., 2015). EEG data were epoched from -200 ms to 400 ms relative to stimulus onset. These epochs were then time-shifted by -22.6 ms to correct for the delay in auditory stimulus delivery, based on the discrepancy between event marker timestamps and the actual onset of the stimulus signal as recorded by the StimTrak system. The epoched signals were then bandpass filtered from 50 to 300 Hz using a finite impulse response (FIR) filter with a lower and upper transition bandwidth of 12.5 Hz and 75 Hz, respectively, and a filter length of 6.6 times the lower transition bandwidth. To estimate signal power at the stimulus frequency for each participant, trials were averaged in the time domain for each tone and target condition. Power spectral densities were calculated separately for the 200 ms pre-stimulus baseline and the 200 ms post-stimulus window of these averaged signals. Power was then expressed in decibels (dB) by taking the logarithm of the ratio between post-stimulus and baseline power, multiplied by 10.

PLV analysis

To assess the consistency of phase-locked neural responses to the stimulus, the phase-locking value (PLV) was calculated using the method described by Lachaux and colleagues (1999) and Zhu and colleagues (2013). FFR epochs were cropped to retain only the post-stimulus segment. PLVs were calculated for each participant and condition using a bootstrapping approach. For each iteration, 400 trials were randomly sampled, and the fast Fourier transform (FFT) was applied to each epoch. Phase values at the stimulus frequency were then extracted by taking the angle of the FFT's complex output. The PLV for each bootstrap draw was calculated by averaging the complex exponentials of these phase values across the 400 trials. This procedure was repeated 1000 times to obtain a stable estimate of PLV for each stimulus condition and participant. To assess significance, a null distribution was generated as in Zhu and colleagues (2013), by computing PLVs from randomly drawn phases 1000 times. Observed PLV values were then statistically compared to the null distribution using a significance level of 0.05, and group comparisons were performed between high- and low-performance participants.

ERP analysis

EEG data preprocessed following the electrophysiological data preprocessing steps described in Experiment 1 were used. The PREP preprocessed data was bandpass filtered from 0.1 to 40 Hz using a FIR filter with a lower and higher transition bandwidth of 0.1 Hz and 10 Hz and a filter length of 6.6 times the lower transition bandwidth. The data were then epoched from -200 ms to 500 ms relative to stimulus presentation onset and then downsampled to 1 kHz. To remove eye and muscle artifacts, independent component analysis (ICA) was applied to the preprocessed EEG data to decompose the signal into 15 independent components (ICs). To identify components reflecting EOG artifacts, each component's time course was correlated with the EOG channels and the resulting correlation coefficients were z-scored. Any IC with a z-score of greater than 1.96 was removed from the EEG signal. Trials were baseline corrected using the 200 ms window prior to stimulus onset. Trials were then automatically rejected using the Autoreject package based on their peak-to-peak amplitudes (Jas et al., 2017). Finally, epochs were averaged in the time domain by condition and target tone to isolate the evoked potential.

Amplitudes for ERP components were calculated by taking the mean amplitude within a time window determined by visual inspection of the time-domain signals and from typical analysis windows used in previous research (Key et al., 2005). The window for the N1 component was 100-150 ms, the window for the P2 component was 150-230 ms (Coch et al., 2005), the window for the N2 component was 250-360 ms (Folstein et al., 2007; Nieuwenhuis et al., 2003) and the window for the P3b component was 360-420 (Lange, 2009; Dien et al., 2010, Polich, 2007). Latencies for each peak were calculated by identifying the time point corresponding to the maximum or the minimum value of the peak in each window. For the small peak or inflection point in the P3b window, the latency of the peak was identified as the time point of the zero-crossing of the first derivative of the discrete time series. Additionally, the time window for the P3b component was shifted to 350-420 ms in Experiment 2 to account for the earlier peak latency.

3.3 Results

Behavioral results

Participants generally performed well at the task with a few participants at ceiling. Overall mean task accuracy was 91% (SD = 6.3%). Participants performed significantly worse on trials in which the middle, 200 Hz tone was the target ($F(2, 88) = 58.7$, $p < 0.01$). Posthoc pairwise comparisons with Bonferroni correction applied indicates that accuracies were significantly lower for the 200 Hz target (M = 85.6%, SD = 9.8%) than the 130 Hz target (M = 95.2%, SD = 4.6%; mean difference = 10.0%, $p < 0.01$) and the 280 Hz target (M = 93.1%, SD = 6.5%; mean difference = 9.6%, $p <0.01$). Accuracies were not significantly different between the 130 Hz and 280 Hz targets (mean difference = 2.1%, $p > 0.1$). A median split was taken to separate participants into high and low task performers (Mdn = 92.3%) (Figure 3.1). An Ordinary Least Squares regression was run between years of musical experience and performance accuracy. The model was weakly statistically significant F(1, 39) = 3.63, $p < 0.1$, $R^2 = 8.9\%$ with years of musical experience significantly predicting scores b = 0.5%, $p < 0.1$.

FFR results

A linear mixed-effects model was fitted using lme4 (Bates et al., 2015) on R to examine the effects of stimulus frequency (130, 200, 280 Hz), target frequency (130, 200, 280 Hz), and group (high vs. low performers) on FFR power (Figures 3.2 and 3.3). Subject was included as a random intercept to account for repeated measures. Post-hoc comparisons were conducted using the emmeans package on R (Lenth, 2025).

The analysis revealed a main effect of stimulus frequency, with higher FFR power at higher stimulus frequencies. Estimated marginal means indicated that responses increased from 130 Hz (M = 2.40 μV, 95% CI [1.08, 3.72]) to 200 Hz (M = 4.37 μV, 95% CI [3.05, 5.69]) and 280 Hz (M = 4.99 μV, 95% CI [3.66, 6.31]). Tukey-corrected pairwise comparisons confirmed that power was significantly greater for 200 Hz compared to 130 Hz ($\Delta = 1.97$ μV, SE = 0.40, t(320) = 4.95, $p < .0001$) and for 280 Hz compared to 130 Hz ($\Delta = 2.59$ μV, SE = 0.40, t(320) = 6.50, $p < .0001$). The difference between 200 Hz and 280 Hz was not significant ($p = .27$).

There were also significant interactions between stimulus frequency and group. Low per-

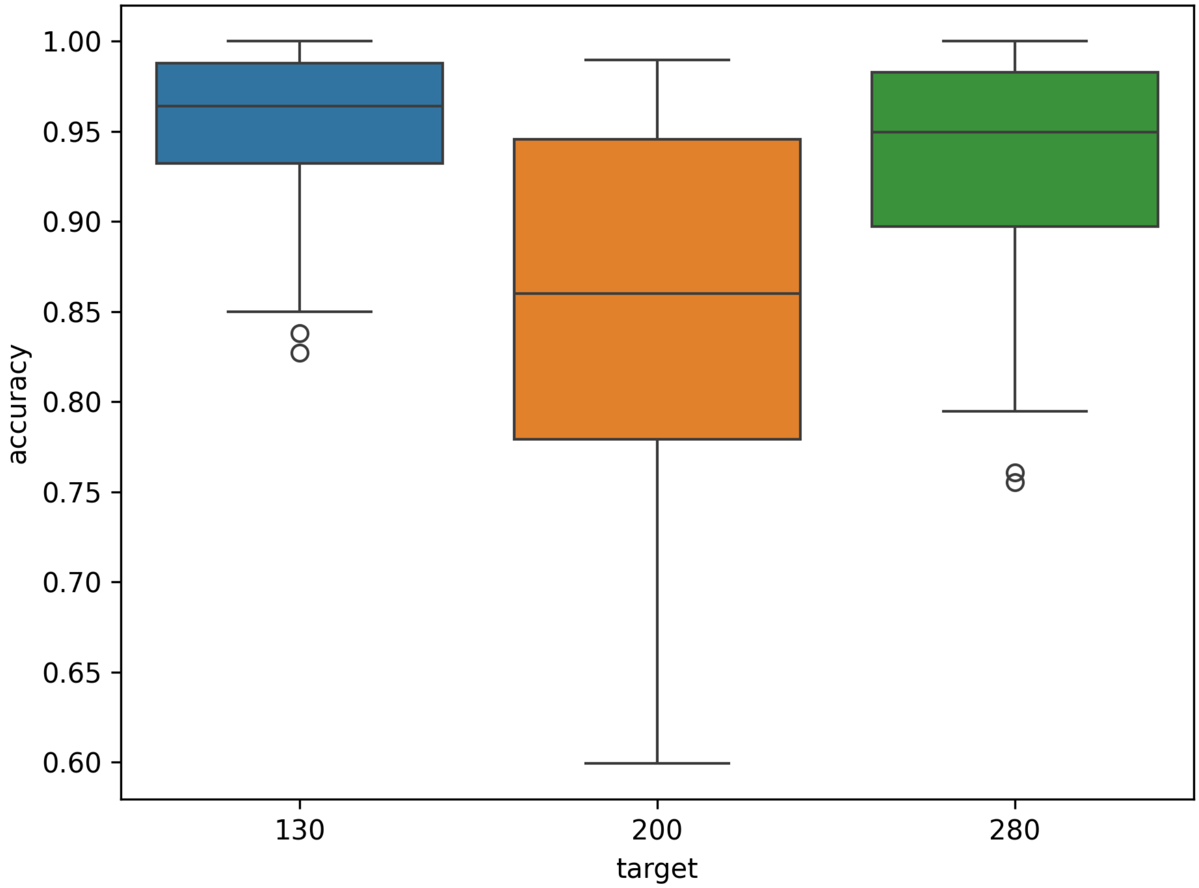


Figure 3.1: Task accuracy was high across all three target conditions. Participants subjectively reported that the middle tone was more difficult, and this is reflected in the accuracy rates.

formers showed consistently higher FFR power than high performers at the higher stimulus frequencies (200 Hz: $\Delta = 2.77$ μV, SE = 1.08, t(59.4) = 2.58, p = .0125; 280 Hz: $\Delta = 3.57$ μV, SE = 1.08, t(59.4) = 3.32, p = .0016), with a nonsignificant trend at 130 Hz ($\Delta = 1.88$ μV, p = .0856). A similar pattern emerged for the target frequency × group interaction: low performers showed greater power across target conditions (130 Hz: $\Delta = 2.22$ μV, p = .0441; 200 Hz: $\Delta = 3.24$ μV, p = .0038; 280 Hz: $\Delta = 2.77$ μV, p = .0125). Finally, the three-way interaction between stimulus, target, and group approached significance (ps .06–.07). Follow-up pairwise contrasts suggested that the group differences were most pronounced when both the stimulus and target were high-frequency tones (e.g., at 280 Hz, high vs. low performers differed by $\Delta = –3.80$ μV, SE = 1.34, t(130) = –2.83, p = .058).

Taken together, these results indicate that FFR power increased with stimulus frequency and was generally higher among low performers than high performers, particularly at the higher stimulus and target frequencies. The three-way interaction suggests that this performance-related effect may be strongest when both stimulus and target frequencies were high, although these contrasts did not reach conventional levels of significance.

PLV results

To verify whether the observed differences in FFR amplitude were due to diminished phase locking, PLV values were calculated for each tone for both low and high-performing participants. PLV values were computed by bootstrapping values from samples of randomly drawn trials, and the null distribution was computed by drawing random phase values. As a majority of the observed PLV values were not normally distributed, their statistical significance was computed by comparing their median to the 90th and 95th percentiles of the null distribution at 0.047 and 0.054 respectively. As the resulting PLV values were verified by the Shapiro-Wilk test to be non-normal, Mann-Whitney U tests were used to compare PLV values for high- and low-performing participants against each other. Phase locking to

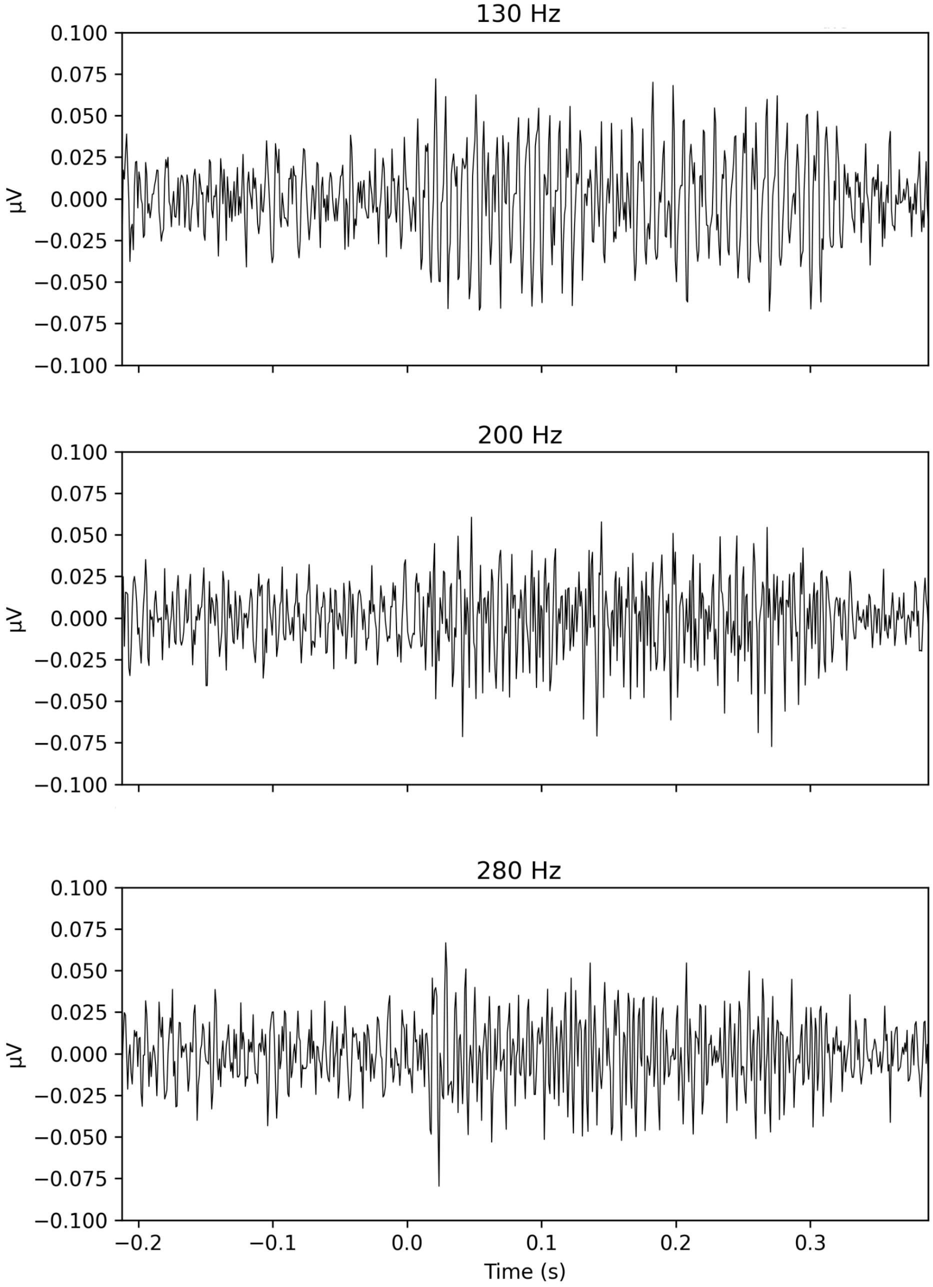


Figure 3.2: Grand average FFRs in the time domain for each tone with stimulus onset is at time 0. The transient brainstem response can be seen in the first 0-50 ms of the waveform, followed by the stable, phase-locked response.

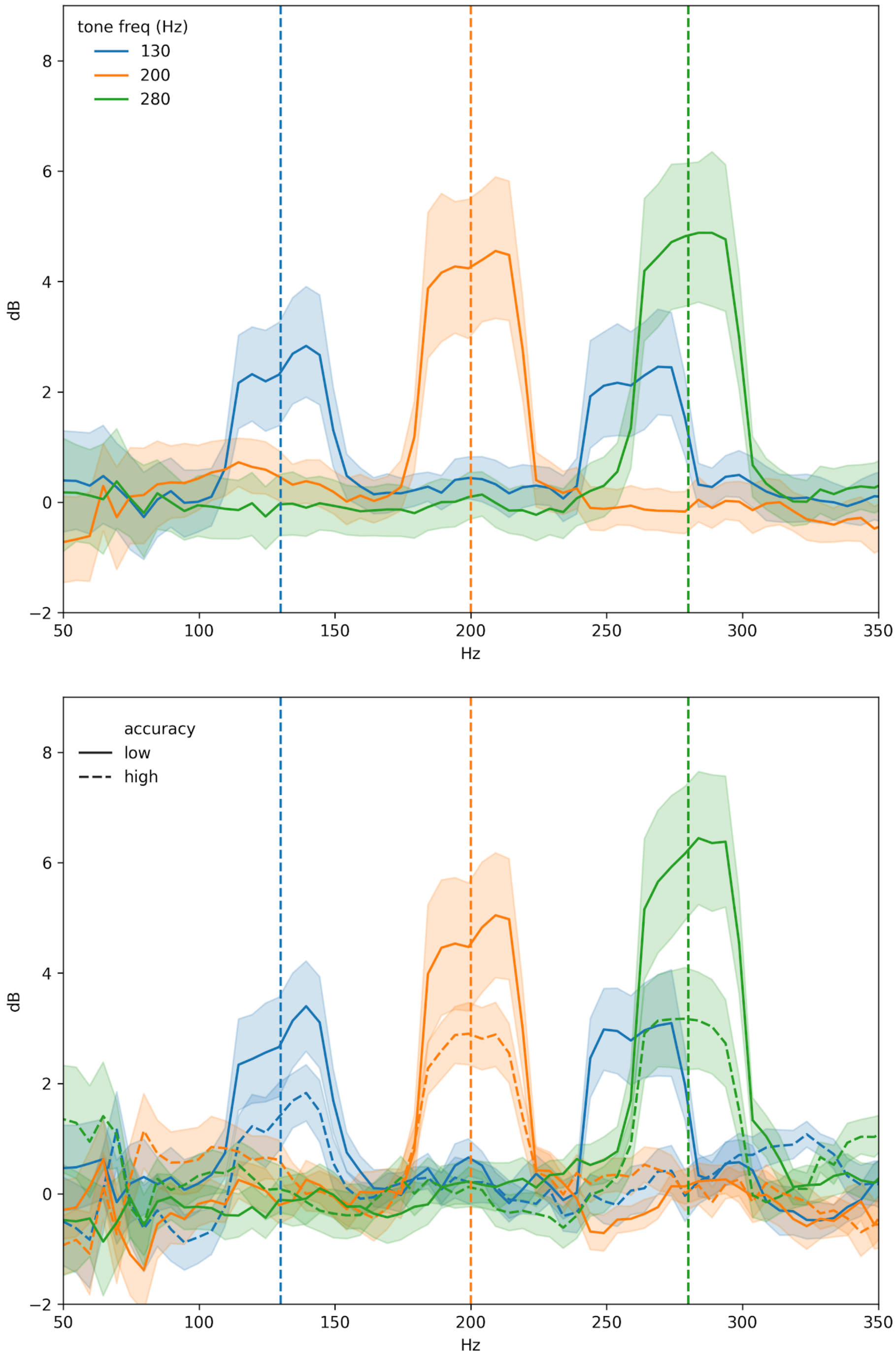


Figure 3.3: FFR amplitudes to each tone. Each trace shows the magnitude spectrum of the evoked response to each tone relative to the baseline period, lines show means and the shaded regions represent the 95% CI. (Top) There is significant power at the pitch of each tone and negligible power at all other frequencies. (Bottom) High task performers had much lower magnitude spectrum amplitudes than low task performers.

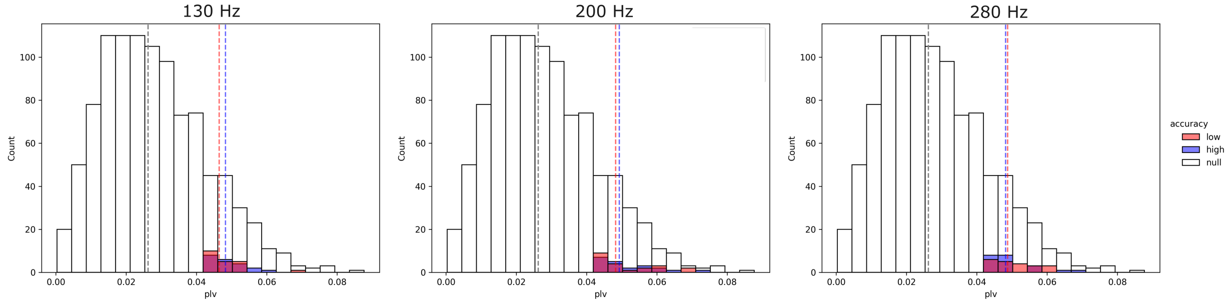


Figure 3.4: Consistency of phase locking to each tone by accuracy group. PLVs across all tones and groups were significantly greater than null, while PLVs between accuracy groups were not significantly different for any of the tones. Dotted lines represent the median value of each distribution.

the 130 Hz tone for high performers (Mdn = 0.048, IQR = 0.005) was significantly greater than null (at $p < 0.1$), while PLVs to the 130 Hz for low performers (Mdn = 0.046, IQR = 0.007) were not ($p > 0.1$). However, phase locking to the 130 Hz tone did not differ between low and high performers ($z = 0.273$, $p > 0.1$). Phase locking to the 200 Hz tone for low performers (Mdn = 0.048, IQR = 0.012) and high performers (0.049, IQR = 0.013) were both significantly greater than null ($p < 0.1$). Phase locking to the 200 Hz tone did not differ between low and high performers ($z = 0.226$, $p > 0.1$). Lastly, phase locking to the 280 Hz tone for low performers (Mdn = 0.048, IQR = 0.009) and high performers (0.048, IQR = 0.005), were both significantly greater than null ($p < 0.1$). Phase locking to the 280 Hz tone did not differ between low and high performers ($z = 0.201$, $p > 0.1$) (Figure 3.4).

ERP results

The amplitudes of multiple ERP components varied significantly as a function of target tone frequency and task performance (Figure 3.5). Omnibus ANOVAs were computed for each component with mean uV as the dependent variable against tone frequency and target frequencies as within-subjects variables and accuracy as a between-participant variable. There was a significant effect of tone frequency ($F(2, 76) = 7.94$, $p < 0.001$) and a significant interaction between tone and target frequency ($F(4, 154) = 6.38$, $p < 0.001$) on the P2 component

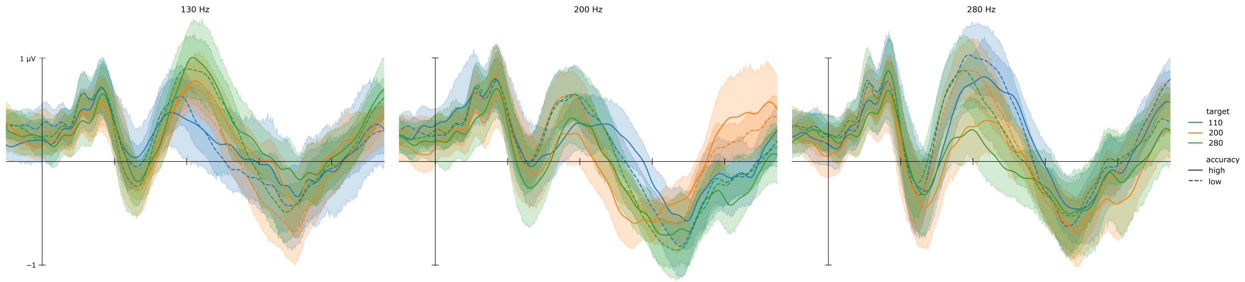


Figure 3.5: Changes in ERP latencies and amplitudes in response to target tones, particularly in the P2 component and the P3b (top). ERP changes to target tones are more pronounced in high-accuracy participants (bottom). Lines represent the mean evoked potential across all participants and the shaded region represents the 95% CI.

at 150-230 msec. Interaction effects suggest that P2 amplitudes were significantly different when a given tone was a target tone. There was a significant effect of tone identity on the N2 component from 230-360 msec ($F(2, 76) = 13.81$, $p < 0.001$), although interaction effects did not reach significance ($F(4, 154) = 1.77$, $p > 0.1$). There was a significant interaction between tone and target frequency on the P3b component at 360-420 msec ($F(4, 154) = 3.23$, $p < 0.05$). There was also a significant effect of accuracy on all of the early components (Figure 3.5, bottom). The main effect of accuracy was significant for the N1 ($F(1, 37) = 10.81$, $p < 0.001$), P2 ($F(1, 37) = 5.16$, $p < 0.05$), and N2 components ($F(1, 37) = 4.72$, $p < 0.05$), but not for the P3b component ($F(1, 37) = 0.01$, $p > 0.1$).

The relationships between target frequency and task performance on the latency of each component were more limited than the amplitude effects. Significant interactions between tone and target, indicating an effect on ERP latencies when a tone is a target tone, were only found for the N2 component ($F(4, 154) = 5.74$, $p < 0.001$). There was a significant influence of task performance on the latency of the N1 ($F(1, 39) = 15.62$, $p < 0.001$) and P2 components ($F(1, 39) = 17.23$, $p < 0.001$). There were no effects of task performance or target identity on the P3b component.

3.4 Discussion

Similar to the previous study, this study compares models of auditory attention by contrasting evidence from electrophysiological responses and analytical methods. As before, the significant associations between task performance and task demands on measures of FFR amplitude and consistency are compared to evaluate gain modulation and dynamic tuning accounts of audition. In addition, the present study employed a variant of the oddball task, which enabled changes in ERP and FFR induced by differences in task performance and task demands to be compared, assessing the extent of top-down modulation on responses in the auditory pathway. As with the previous experiment, FFR amplitudes did not vary significantly with task demand or interactions between task demands, manipulated by target tone frequency, and task performance. In an interesting deviation from the previous experiment, however, lower FFR amplitudes were observed in high-accuracy performers in the task. There were no main effects or interaction effects of task performance or task demands on PLVs. Simultaneous measures of ERPs during FFR recording show significant increases in peak amplitudes for the P2 and P3b components in response to target tones, with an interaction with task performance indicating a stronger effect in high performers. There were limited latency effects except for a main effect of target tone on the N2 component.

Disentangling effort and efficiency

The observed decrease in FFR amplitude with increased task performance is consistent with predictions made by dynamic tuning models of auditory attention and expertise. These models propose that attentional enhancement and auditory training do not simply amplify all neural responses but instead sharpen the tuning of neural populations to behaviorally relevant features (e.g., Fritz et al., 2007; Shamma & Fritz, 2014). This sharpening entails increased selectivity, wherein neural responses to target features become more temporally precise and narrowly focused, while activity unrelated to the task is suppressed. This finding is consistent with data from Chelazzi and colleagues (1993), which show a decline in the firing rate of cells not tuned to the properties of the attended stimulus as the stimulus becomes more familiar. This finding is also consistent with evidence from Schwartz and David (2018), which shows decreased firing rates in neurons tuned to task-irrelevant information. In this framework, the reduced FFR amplitude in high performers may reflect a more refined and efficient allocation of neural resources, suggesting that fewer, more selectively tuned neurons are phase locking to the stimulus, thereby maintaining response fidelity without increasing overall neural recruitment and, consequently, energy expenditure.

The lack of significant differences in PLV between high and low performers further supports this interpretation. While PLV indexes the consistency of phase-locked responses, amplitude reflects the overall strength or extent of neural recruitment. Dynamic tuning models predict that neural precision may increase without a corresponding increase in response magnitude, especially in individuals with enhanced auditory attention or expertise (Snyder & Alain, 2007; Fritz et al., 2007). In other words, high performers may achieve more efficient encoding by selectively engaging a smaller, more functionally tuned subpopulation of neurons. These results contrast with gain modulation theories, which would predict increased amplitudes and PLVs under greater attentional engagement. Instead, our findings suggest that expertise-related neural plasticity leads to a qualitative reorganization of auditory processing, marked by enhanced selectivity rather than increased overall excitability.

While the present study found a decrease in FFR amplitudes in experts, the previous study reported an increase in FFR amplitudes. Although both experiments used variants of the oddball tasks to manipulate selective attention, differences in the details of their experimental design may have contributed to the observed differences. In the first task, target tones were rare tones that were slightly higher or lower in pitch than the standards. In comparison, trials in the second task consisted of randomized sequences of tones selected from three possible candidates. A different tone was selected to be the target tone in each trial, such that target tones appeared at an equal probability to standard tones. Moreover, while participants were rewarded points based on the accuracy of their answers in both tasks, the criterion for earning a point in the first experiment was much stricter. In the current study participants could earn a point if their answers were within two counts of the correct answer. These experiment design differences resulted in the second task being much easier than the first, as reflected in the higher average accuracy scores. One possible factor contributing to observed differences in FFR effects may therefore be the difficulty of the task and the amount of effort expended by listeners.

The Framework for Understanding Effortful Listening (FUEL) by Pichora-Fuller and colleagues (2016) discusses the variable effects of task difficulty, cognitive demand, and motivation in jointly determining listening effort and cognitive resource allocation. According to FUEL, increased task difficulty heightens listening effort by drawing on working memory, executive control, and attentional resources. As a result, high performers in difficult tasks may exhibit increased FFR amplitudes not because they represent the stimulus more efficiently, but because they are investing cognitive resources to enhance perceptual encoding and maintain accurate performance. This is consistent with studies that report greater pupil dilation under increased task loads (Zhang et al., 2021; Kramer et al., 1997). Conversely, in easier tasks, high performers may rely on more efficient, selective processing strategies that reduce overall neural activation, leading to reduced FFR amplitudes despite strong behavioral performance. These contrasting patterns emphasize an important distinction: stronger neural responses do not necessarily reflect better encoding, and weaker responses do not necessarily indicate worse encoding. Instead, FFR amplitude may reflect the degree of cognitive effort rather than the quality of representation. This distinction highlights a critical gap in the attention literature, where effective attention is often conflated with effortful attention. Future studies of attentional modulation in auditory neuroscience should consider effort and efficiency as distinct constructs, particularly when interpreting neurophysiological measures, such as FFR amplitude.

If FFR amplitude reflects the degree of cognitive effort rather than the precision or quality of stimulus representation, this challenges a central assumption of gain modulation theories of auditory attention. Gain-based models (e.g., Hillyard et al., 1998; Zatorre et al., 2004, 2012) posit that attention enhances sensory encoding by increasing the magnitude of neural responses to relevant stimuli. Under this framework, stronger FFR amplitudes are typically interpreted as evidence of improved auditory representation under attentive or expert conditions. However, if amplitude increases primarily signal greater effort or resource allocation, then high FFR amplitudes might instead reflect the neural cost of maintaining performance under challenging conditions, rather than a faithful or precise encoding of the stimulus. This interpretation challenges the notion that amplitude gain is a reliable proxy for representational quality in all cases.

In contrast, dynamic tuning models (e.g., Fritz et al., 2007; Schreiner & Polley, 2014) emphasize that attention operates by sharpening receptive fields and reducing responses to irrelevant information, thereby increasing neural selectivity rather than overall response strength. From this perspective, the decreased FFR amplitude in high performers on easier tasks may reflect more efficient and selective encoding, consistent with refined neural tuning that suppresses irrelevant or redundant information. If PLV remains high while amplitude decreases, this supports the idea that fewer neurons are recruited, but those that are involved phase lock more consistently to the stimulus. Thus, viewing FFR amplitude as a proxy for effort rather than fidelity is congruent with dynamic tuning models, which prioritize the precision of neural engagement over the magnitude of the response. This reconceptualization invites a more nuanced interpretation of FFR data, particularly in studies of attention and expertise, and underscores the need to jointly consider both phase-based and amplitude-based metrics.

These results highlight how expertise can allow individuals to flexibly allocate cognitive and attentional resources depending on task demands. Rather than producing a uniformly stronger FFR, expertise may enable listeners to strategically engage different neural mechanisms depending on the complexity of the listening environment. In difficult tasks, high performers may increase gain to enhance perceptual clarity, while in easier tasks they may rely on dynamic tuning mechanisms to suppress irrelevant input and streamline processing. This flexibility suggests that gain modulation and dynamic tuning are not mutually exclusive but may instead represent complementary strategies that are differentially recruited depending on task difficulty and cognitive load. Expertise, shaped by long-term auditory experience, may facilitate more efficient switching between these modes, thereby optimizing neural resource allocation in real-time. Together, these findings argue for a more integrative framework in which attentional effects on the FFR reflect both the amount and efficiency of neural engagement, modulated by the interplay between task difficulty, effort, and listener expertise.

Limits of top-down attention on early auditory encoding

The current experiment revealed robust enhancements in ERP components such as the P2 and P3b in response to target tones, consistent with the established literature on attention and target detection (Luck, 2011; Polich, 2007; Näätänen et al., 1992). ERPs are late, large-amplitude potentials that are generally accepted to reflect higher-order cognitive processes. The P2 component is often associated with stimulus evaluation and auditory categorization (Luck & Hillyard, 1994; Näätänen & Winkler, 1999). The P3b, sometimes known as the P3b, is a late positive component that peaks around 300-500 ms after stimulus onset. The P3b is associated with attentional resource allocation and context updating (Sutton et al., 1965; Polich, 2007), and auditory temporal attention and sequential expectancy (Dien et al., 2010; Jentzsch and Sommer, 2001; Sommer et al., 1999; Lange et al., 2006). Crucially, these ERP effects were significantly amplified in participants with higher auditory task performance, suggesting more efficient or more engaged processing of task-relevant auditory stimuli at later stages of the neural hierarchy. The sensitivity of these ERP components to both task demands and task performance reinforces the idea that the task manipulation effectively engages attentional systems at a cognitive level.

In contrast, the absence of corresponding effects on FFR amplitude or phase locking to target tones suggests that early auditory encoding, as reflected in the FFR, may be relatively insensitive to rapid, trial-by-trial shifts in stimulus relevance. This dissociation between early and late neural markers strengthens the conclusion from earlier experiments that immediate, top-down effects of task demands exert only a limited influence on the FFR. If attention modulated early sensory encoding in a robust and flexible manner, we would expect to see enhanced FFRs to target tones relative to non-targets, especially in high-performing individuals. Instead, the presence of strong task effects in ERPs but not in the FFR implies that the influence of attention on auditory encoding may become more prominent only after initial subcortical processing, in line with models proposing a cortical locus for attentional gain (Woldorff & Hillyard, 1991; Bidelman & Momtaz, 2021). These findings further support the view that the FFR primarily reflects bottom-up stimulus-driven processes, with limited dynamic modulation by short-term attentional goals.

Lastly, the P2 has also been shown to decrease with the level of attentiveness and it has been suggested that the modulation of the N1-P2 complex may index the ease of distinguishing between relevant and irrelevant information (Crowley & Colrain, 2004; Näätänen, 1990). The interaction between target identity and task performance in increasing P2 amplitudes strengthens the claim that high task performers expended less effort during the task. The decreased FFR amplitudes in conjunction with the increased P2 amplitudes in high performers indicate that the FFR may index effort, and that increased FFR amplitude does not necessarily reflect enhanced stimulus encoding.

Conclusion

Taken together, these results highlight a crucial distinction between the neural mechanisms underlying cognitive effort and the quality of auditory encoding. Increased FFR amplitudes observed in more difficult tasks may reflect greater cognitive resource allocation rather than enhanced stimulus representation, consistent with the FUEL framework and gain modulation theories. In contrast, decreased FFR amplitudes in easier tasks suggest that efficient, selective neural encoding may not always require heightened gain, aligning with predictions from dynamic tuning models. The absence of FFR modulation by target tones, despite robust P2 and P3b ERP effects, further supports the view that early auditory encoding is relatively stable and less sensitive to immediate task demands. In other words, these results suggest that top-down attentional effects may be apparent only in later, cortical processing stages. Overall, these results imply that gain and tuning mechanisms may co-occur, with gain reflecting transient effort-related boosts in neural responsiveness, and dynamic tuning reflecting longer-term optimization of selective encoding. This flexibility implies that expert listeners dynamically adjust the neural gain and selectivity of their responses to suit the task context. Future models of auditory attention may benefit from integrating both mechanisms, allowing for shifts between effortful listening and efficient neural selectivity depending on task demands, expertise, and cognitive load.

CHAPTER 4

EXPERIMENT 3: DIVERGENT MECHANISMS OF ATTENTION AND INATTENTION IN EARLY AUDITORY ENCODING

4.1 Introduction

Continuing the investigation into the interactions between expertise, effort, and attention in the mechanisms of auditory encoding, the present study extends the previous findings by incorporating a passive listening component into the experiment. The first study in this dissertation found higher FFR amplitudes in high performers during an intermodal attention task, with no effects of task modality or interaction effects between task demands and task performance. The second study in this dissertation, however, with an easier task and a more lenient scoring system for incentivizing participants, found decreased FFR amplitudes in high performers. Again, this study found no effects of task demands, manipulated by target tone identity, or any interaction effects. Results from these studies suggest the need for a closer examination of the differences between increased cognitive resource allocation and subsequent gain modulation during effortful listening, and the more efficient encoding associated with suppressed activity in high performers under easier task demands. Comparing the interactions between task performance and task demands during active and passive listening will help develop an understanding of how gain modulation and dynamic tuning mechanisms are differentially recruited. Moreover, past experiments on the effects of expertise on the FFR have recorded FFRs exclusively under passive listening conditions, adding a passive listening component will ground the current results in that body of literature.

Gain modulation and dynamic tuning are two prominent theoretical accounts of how attention modulates sensory processing, including auditory encoding. Gain modulation has been found in late, low-frequency, large population responses measured by methods such as fMRI and ERP. Models of perceptual processing based on this data posit attention enhances perception by amplifying neural responses to relevant stimuli, effectively increasing the gain or magnitude of sensory signals (Hillyard et al., 1998; Zatorre et al., 2002; Pantev et al., 1998). According to this view, attending to a stimulus, whether through voluntary focus or expertise, should lead to stronger evoked responses as more neurons are recruited or existing responses are amplified. In the auditory domain, gain modulation predicts that attention and task engagement should increase the amplitude of neural responses such as the frequency-following response (FFR), and that these increases should be particularly evident during active listening when cognitive resources are directed toward monitoring the stimulus stream. In contrast, passive listening, where attention is not explicitly engaged, should result in lower FFR amplitudes due to reduced top-down modulation.

Dynamic tuning models, by contrast, emphasize the refinement of sensory representations through selective attention rather than simple amplification. These models propose that attention sharpens neural tuning by enhancing responses to task-relevant features and suppressing irrelevant information (Fritz et al., 2007; Schreiner & Polley, 2014). Rather than increasing response magnitude across the board, dynamic tuning predicts more efficient and selective neural activity, engaging a smaller population of neurons that are more precisely tuned to the attended feature. This can manifest as decreased amplitude in measures like the FFR if irrelevant inputs are being actively suppressed. Under passive listening conditions, when top-down selection is minimal, dynamic tuning predicts broader, less selective activation and potentially higher FFR amplitudes due to a lack of suppression. Thus, while gain models anticipate attenuated responses during passive states, dynamic tuning models may predict enhanced or less focused responses in the absence of attentional constraints. These divergent predictions make the active versus passive contrast a powerful contrast for disentangling the contributions of gain versus tuning mechanisms in auditory attention.

Despite the theoretical differences between gain modulation and dynamic tuning models, the limitations in current FFR research are ambiguous as to whether attention and expertise increase the gain or selectivity of FFRs. Few studies have included within-subject comparisons of passive and active listening conditions, which are essential for testing the contrasting predictions made by gain and tuning theories. The majority of FFR research on expertise or attentional effects has measured electrophysiological responses during passive listening (e.g., Parbery-Clark et al., 2009; Musacchia et al., 2008; Lehmann & Schönwiesner, 2014; Omote et al., 2017), with other predictors measured separately, making it difficult to isolate the role of attentional state from between-subject confounds.

Moreover, relatively few studies incorporate behavioral performance measures such as task accuracy as predictors of neural response variability. This omission limits the ability to distinguish between the effects of effective attention and effortful attention, a distinction that is increasingly emphasized in models of listening effort and auditory cognition (Pichora-Fuller et al., 2016; Peelle, 2018). Findings from the past two studies in this dissertation suggest that effort and efficient encoding can vary independently and may recruit different neural mechanisms. While increased FFR amplitudes in high performers during a difficult task may reflect heightened cognitive resource allocation and gain modulation, decreased FFR amplitude in high performers during an easier task likely reflects more efficient, selective encoding with less neural expenditure. These contrasting patterns suggest that interpreting FFR amplitude as a straightforward index of encoding accuracy or fidelity can be misleading. To address this, the current study introduces an analysis separating participants based on their behavioral trajectories across the experiment: consistent high performers, (indicative of effective attention with low effort), "improvers" who show marked gains in performance over time (effective attention with high effort), and "decliners" whose performance worsens or remains poor (ineffective attention and low effort). By examining FFRs and ERPs across these subgroups, the study aims to differentiate neural signatures of attentional control from those of resource mobilization, providing a more nuanced understanding of how expertise and cognitive effort interact in shaping early auditory encoding.

The primary objective of this study is to replicate the main effects of reduced FFR amplitudes under active listening conditions in high performers. As with the previous experiments, since many of the high task performers are at the task ceiling and task performance has been found to significantly correlate with musical experience, task performance will be used as a measure of expertise. The second aim of this study will therefore be to investigate the relationship between expertise and task demands in active versus passive listening tasks. Findings from the prior chapters highlight the importance of distinguishing between increased cognitive resource allocation and the associated gain modulation observed during effortful listening, as well as the more efficient, selective encoding linked to suppressed neural activity in high performers under less demanding tasks. Examining how task performance and task demands interact across active and passive listening conditions can shed light on when and how gain modulation versus dynamic tuning mechanisms are engaged. Lastly, as with the previous two experiments, this experiment investigates the interactions between task demands and task performance to probe whether the lack of task demand and selective attention effects in previous studies may be attributed to individual differences in effort or the effectiveness of the attention manipulation.

4.2 Materials and methods

A total of 28 new participants were recruited for this experiment from the University of Chicago community (18 women, 10 men). These participants had not previously completed Experiment 1 or 2. All participants had normal speech and hearing as verified with a Welch Allyn otoscope. Data from two participants were excluded because fewer than 80% of FFR trials were retained following trial rejection using a peak-to-peak rejection threshold of 35 mV on the PREP preprocessed data (Bigdely-Shamlo et al., 2015). Participants signed an informed consent form approved by the Institutional Review Board at the University of Chicago prior to their participation. Participants were compensated with either course credit or \$20/hr in cash.

Stimuli and task

Task setup and stimulus presentation were identical to those in Experiment 2. The stimuli used in this task were three pure sine-wave tones at 130, 200, and 280 Hz. Tones onsets and offsets were smoothly ramped with a 5 ms Hanning window. The tones were presented at 65–70 dB SPL over insert earphones (Etymotic ER3C). The tones were 300 ms in duration, with a jittered interstimulus interval ranging from 200 to 250 ms. The length of sequences varied uniformly between 42 and 48 tones. One of the three tones was randomly selected to be the target tone at the start of each trial. Before each trial, participants were allowed to listen to the target tone as many times as they wished. Once the trial started, participants were asked to count the number of times the target tone played and report their answer at the end of the trial. Tone sequences were pseudo-randomized such that the first tone was never a target tone, and no tone would play more than three consecutive times. As with Experiment 2, participants were given one 'point' if they answered correctly, or if their guess came within 2 tones of the actual target count. Participants were informed they would be rewarded with up to an additional \$10 in cash depending on task performance, although in practice, all participants were rewarded with the bonus. Participants completed 3 blocks of 18 trials each, each lasting approximately 12-14 minutes. The entire task lasted approximately 1 hour, with 800-900 sweeps per tone per condition.

Participants also completed a passive listening task, in which participants sat and passively listened to two blocks of tones. The tones and ISI were identical to those used in the active task. Each block contained one long, randomized sequence of 1200 tones, lasting 10.5 minutes. Participants were simply instructed to sit and rest while the tones played; the block was referred to as a "resting" or "recording" block, with no explicit instructions about listening included. Passive and active task orders were counterbalanced between participants such that all odd-numbered participants received the active task first, and all even-numbered participants received the passive task first.

As participants were asked to silently count the number of targets in each trial, behavioral performance for each trial was calculated as 1- the ratio between the error and the actual number of targets. The overall accuracy for each participant was calculated as their mean accuracy for the first 18 trials of each block. The median accuracy was calculated and participants with accuracies below the median were labeled as low performers, while participants with accuracies above the median were labeled as high performers.

Electrophysiological data acquisition and preprocessing

All EEG recording and preprocessing procedures followed those implemented in Experiments 1 and 2. Prior to the placement of the EEG cap, participants were briefed on the task and EEG procedures. The head circumference of each participant was measured to ensure an appropriate fit for the actiCAP slim 64-channel EEG net (actiCAP, Brain Products GmbH, Germany). Participants were seated in a well-lit, sound-attenuated, and electromagnetically shielded room and instructed to attend to the auditory stimuli while minimizing eye blinks and other sources of movement artifacts. Upon completion of the experiment, the precise locations of all electrodes were recorded using CapTrak (Brain Products GmbH, Germany). EEG signals were recorded at a sampling rate of 5 kHz using an actiCHamp Plus amplifier (Brain Products GmbH, Germany). To monitor ocular movements, two electrodes (corresponding to AF7 and AF8 in the actiCAP layout) were placed below the outer canthi of the left and right eyes, serving as electrooculography (EOG) channels. Preprocessing of the EEG data was performed using custom Python scripts in conjunction with the MNE-Python library version 1.5.0 (Gramfort et al., 2013). The eye channels (AF7 and AF8) were re-referenced to Fp1 and Fp2, respectively, forming bipolar EOG channels by referencing electrodes below the eyes to those above the contralateral eyes. The automated PREP pipeline was used to apply a 60 Hz notch filter, detect and exclude noisy channels using threshold-based artifact rejection, and re-reference the EEG data to the common average. Participant-specific preprocessing reports, including metrics such as retained trial counts, removed Independent Components, interpolated electrodes, and visual diagnostics for quality control, are publicly available with the dataset on OpenNeuro. For both experiments, timing calibration was conducted prior to data collection, revealing an average delay of 26.8 ms with a jitter of 0.69 ms in event timing.

FFR analysis

The analyses of FFR amplitude and PLV followed the same procedures used in Experiment 1. The common average reference (CAR) applied during the PREP preprocessing pipeline was retained to mitigate potential bias toward peripheral auditory generators, such as the cochlear microphonic or auditory nerve, and to avoid inflating estimates of absolute FFR amplitude (Bidelman et al., 2015). Continuous EEG data were segmented into epochs ranging from –200 ms to 400 ms relative to stimulus onset. To correct for auditory presentation delay, each epoch was time-shifted by 26.8 ms, based on the measured discrepancy between event marker timestamps and the actual stimulus onset, as recorded by the StimTrak system. The resulting epochs were then bandpass filtered from 50 to 300 Hz using a finite impulse response (FIR) filter with a lower transition bandwidth of 12.5 Hz, an upper transition bandwidth of 75 Hz, and a filter length corresponding to 6.6 times the lower transition bandwidth. To quantify signal power at the stimulus frequency, trials were first averaged in the time domain for each tone and target condition. Power spectral densities were computed for the pre-stimulus baseline (–200 to 0 ms) and the first 200 ms following stimulus onset. Power was expressed in decibels (dB) by taking the logarithm of the ratio between post-stimulus and baseline power, multiplied by 10.

PLV analysis

The phase-locking value (PLV) was computed to evaluate the consistency of phase locked neural responses to the periodic auditory stimuli, following methods described by Lachaux and colleagues (1999) and Zhu and colleagues (2013). For this analysis, each FFR epoch was cropped to include only the post-stimulus segment. PLV estimates were generated for each participant and condition using a bootstrapping procedure. In each iteration, 400 trials were randomly sampled, and the fast Fourier transform (FFT) was applied to each epoch. Phase information at the stimulus frequency was extracted by taking the angle of the complex-valued FFT output. The PLV for that iteration was calculated by averaging the complex exponentials of these phase angles across all 400 trials. This procedure was repeated 1000 times to produce stable PLV estimates for each participant and condition. To assess statistical significance, a null distribution was generated by computing PLVs from 1000 random draws of phase values, following the procedure described in Zhu et and colleagues (2013). Observed PLV values were compared to the simulated PLVs using an alpha of 0.05. Between-group comparisons were conducted to evaluate differences in PLV between high- and low-performance participants and across conditions.

4.3 Results

The results are presented in two parts. First, behavioral and neural responses are reported for the full sample. Then, follow-up analyses that examine how these patterns differ across participant subgroups defined by task performance trajectories are presented.

Behavioral results

As with Experiment 2, participants found this task relatively easy, and a few participants were at ceiling. The overall mean accuracy was 91% (SD = 6.4%). However, performance varied significantly depending on which tone served as the target (F(2, 56) = 25.7, $p < 0.01$). Post hoc pairwise comparisons with Bonferroni correction revealed that accuracy was significantly lower for trials with the 200 Hz tone as the target (M = 85.1%, SD = 10.6%) compared to both the 130 Hz target (M = 94.6%, SD = 5.2%; mean difference = 9.5%, $p < 0.01$) and the 280 Hz target (M = 92.9%, SD = 6.2%; mean difference = 7.8%, $p < 0.01$). There was no significant difference in accuracy between the 130 Hz and 280 Hz targets (mean difference = 1.7%, $p > 0.1$). To facilitate subsequent analyses, a median split was used to divide participants into high and low task performers (median = 91.3%) (Figure 4.1). Additionally, an simple linear regression was run revealing a statistically significant relationship between years playing an instrument and task performance (b = 0.6%, $p < 0.01$).

FFR results

A linear mixed-effects model with random intercepts was fitted to examine the effects of task condition (passive vs. active), task order (passive-first vs. active-first), tone frequency (130, 200, 280 Hz), and task accuracy (high vs. low performers) on FFR amplitudes (Figure 4.2 and 4.3). Subject was included as a random intercept to account for repeated measures. Analyses were conducted with the lme4 (Bates et al., 2015) and emmeans packages on R (Lenth, 2025).

The analysis revealed a main effect of tone frequency. Estimated marginal means indicated that amplitudes were highest for the 200 Hz tone (M = 6.26 μV, 95% CI [4.94, 7.57]) compared to both 130 Hz (M = 2.82 μV, 95% CI [1.51, 4.13]) and 280 Hz (M = 3.14 μV, 95% CI [1.83, 4.45]). Tukey-corrected pairwise comparisons confirmed that amplitudes were significantly greater for 200 Hz relative to 130 Hz ($\Delta = 3.43$ μV, SE = 0.43, t(266) = 8.06, $p < .0001$) and 280 Hz ($\Delta = 3.12$ μV, SE = 0.43, t(266) = 7.32, $p < .0001$). The difference between 130 Hz and 280 Hz was not significant ($p = .74$).

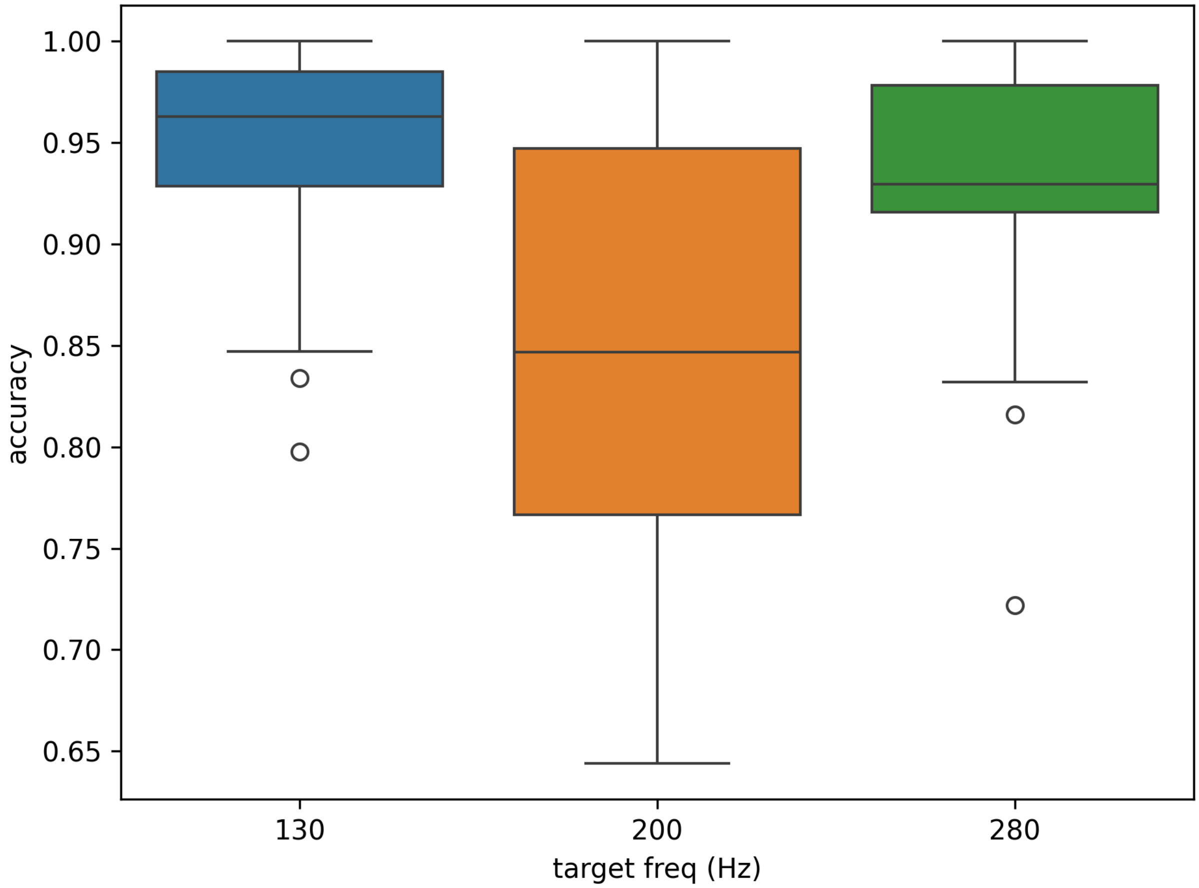


Figure 4.1: Behavioral performance depended on target tone frequency. As with Experiment 1, accuracy was lowest for the middle tone, while there was no difference in accuracy between the lowest and highest target tones.

A main effect of task condition was observed at trend level (p = .112), but did not reach significance when task order was included. Task order itself was not significant (p = .93). However, a modest interaction between task order and tone frequency emerged, driven by a significant effect for the 280 Hz tone, b = –2.62, SE = 1.21, t(266) = –2.17, p = .031. Post hoc comparisons indicated that amplitudes to the 280 Hz tone tended to be lower when the passive task was presented second (M = 2.89 μV) compared to when it was presented first (M = 3.38 μV), although the direct contrast was not statistically significant (p = .646). In addition, there was a tone × accuracy interaction, with low performers showing greater amplitudes than high performers at 200 Hz (M = 7.41 vs. 5.10 μV, Δ = –2.31 μV, SE = 1.05, t(38.5) = –2.20, p = .034). Group differences at 130 Hz and 280 Hz were not significant (both ps > .16). No higher-order interactions reached significance (all ps > .30).

Together, these results replicate the robust effect of tone frequency, with particularly strong responses to 200 Hz tones, and indicate that task accuracy influences amplitudes selectively at 200 Hz. Task order interacted with tone frequency only weakly, suggesting that presentation order may have a minor influence on responses to the highest-frequency tone.

To further examine whether target tone frequency influenced FFR amplitudes, a separate linear mixed-effects model was fitted on data from the active task only. This separate analysis was necessary because target frequency was nested within task condition and could not be included as a factor in the previous omnibus model. The model included fixed effects of tone frequency (130, 200, 280 Hz), target frequency (130, 200, 280 Hz), task accuracy (high vs. low performers), and their interactions, with subject entered as a random intercept. As before, results revealed a robust main effect of tone frequency, with greater amplitudes for 200 Hz tones (M = 4.29 μV, 95% CI [3.14, 5.43]) compared to both 130 Hz (M = 1.80 μV, 95% CI [0.65, 2.94]) and 280 Hz tones (M = 1.57 μV, 95% CI [0.43, 2.72]). Tukey-corrected pairwise comparisons confirmed significantly larger amplitudes for 200 Hz relative to 130 Hz

($\Delta = 2.49$ μV, SE = 0.41, t(192) = 6.14, $p < .0001$) and 280 Hz ($\Delta = 2.72$ μV, SE = 0.41, t(192) = 6.69, $p < .0001$). There was no significant difference between 130 Hz and 280 Hz (p = .84). Crucially, there was no main effect of target frequency (all ps > .36), and no interactions between target frequency and tone or accuracy reached significance (all ps > .23). Thus, in the active task, FFR amplitudes were strongly determined by the eliciting tone frequency but were unaffected by the frequency of the subsequent target stimulus.

PLV results

PLVs were calculated as a measure of phase locking consistency. PLVs were calculated for each tone for both low and high-performing participants and under both passive and active listening tasks. Mean PLVs for each participant were calculated by randomly sampling trials and computing PLV values for each draw. The null distribution was computed by drawing random phases and calculating their PLVs 1000 times. Given the small sample size and the fact that most of the observed PLV values were not normally distributed, their statistical significance was determined by comparing their median to the 95th percentile of the null distribution, which was 0.055. Mann-Whitney U tests were used to compare condition groups against each other. Phase locking to the 130 Hz tone for low performers (Mdn = 0.071, IQR = 0.031) and high performers (0.061, IQR = 0.008), were both significantly greater than null (Mdn = 0.026, IQR = 0.021, both $p < 0.001$). Additionally, phase locking to the 130 Hz tone did not differ between low and high performers (z = 0.55, $p > 0.1$). Similarly, phase locking to the 200 Hz tone for low performers (Mdn = 0.101, IQR = 0.067) and high performers (Mdn = 0.081, IQR = 0.044) were both significantly greater than null ($p < 0.05$). Phase locking to the 200 Hz tone was not different between low and high performers (z = 0.55, $p > 0.1$). Lastly, phase locking to the 280 Hz tone for low performers (Mdn = 0.065, IQR = 0.008) and high performers (Mdn = 0.060, IQR = 0.013) were both significantly greater than null ($p < 0.05$). Phase locking to the 280 Hz tone was not different between low and

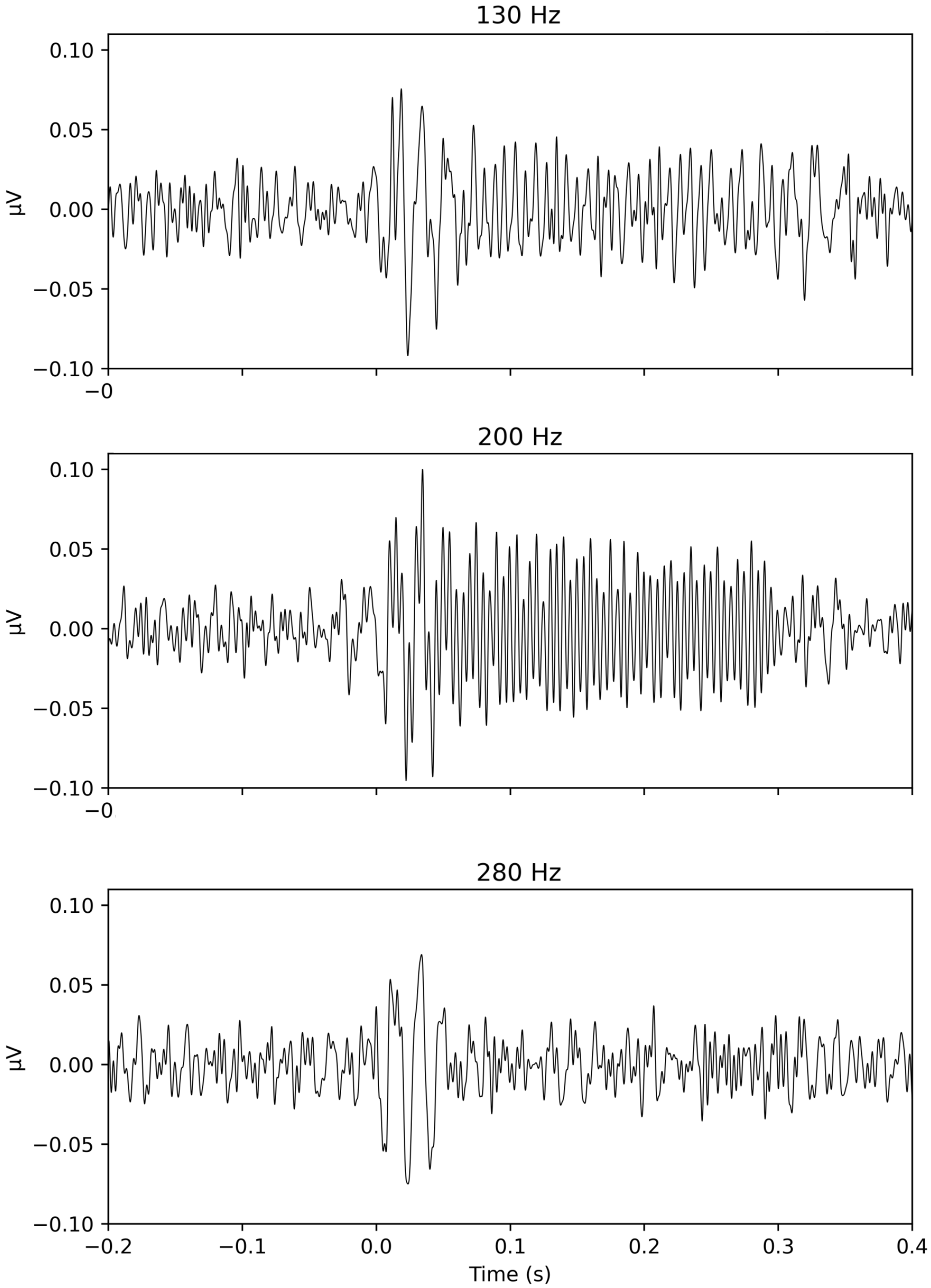


Figure 4.2: Grand average FFRs in the time domain for each tone. Stimulus onset is at time 0. The transient brainstem response can be seen in the first 0-50 ms of the waveform, followed by the stable, phase locked response.

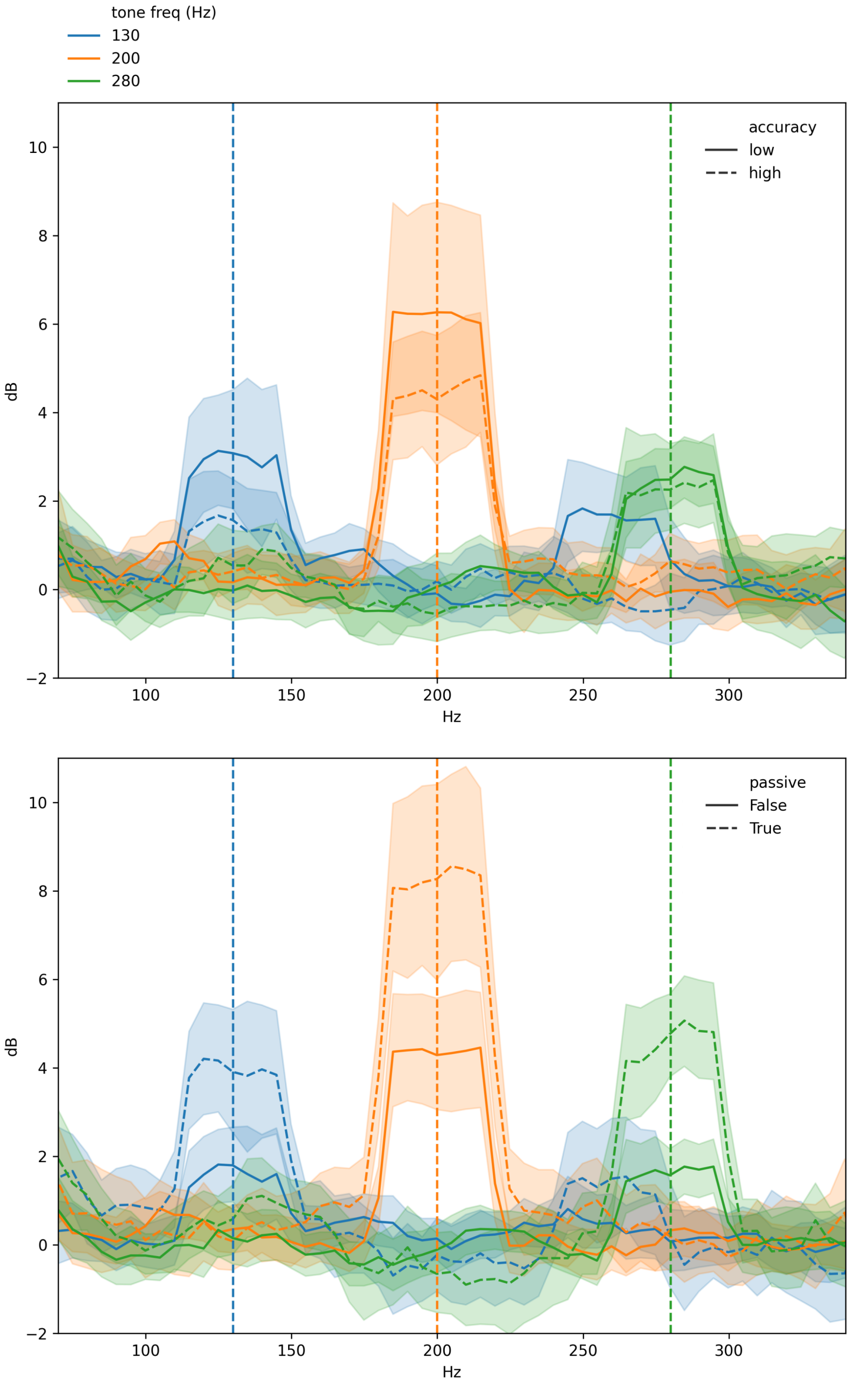


Figure 4.3: Main effects of tone frequency, task condition, and task performance on FFR amplitudes. Lines indicate the mean amplitude of the FFR signal in dB for a given condition. Shaded regions represent the 95% CI of the mean. FFR amplitudes are attenuated in high task performers (top). FFR amplitudes are also attenuated during the active task (bottom).

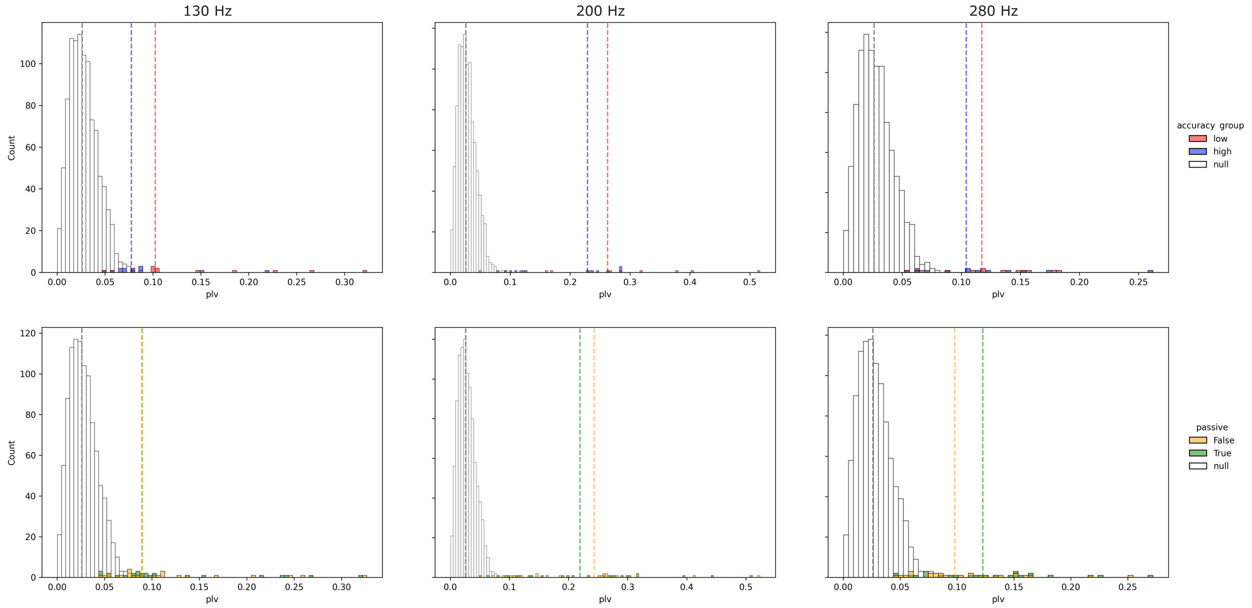


Figure 4.4: PLVs in both high and low task performers were higher than chance for all tones. There was no difference in PLVs in high and low task performers for any of the tones (top). PLVs during both passive and active listening tasks were greater than chance for all tones. There was no difference in PLVs during active and passive tasks for any of the tones (bottom).

high performers ($z = 0.57$, $p > 0.1$) (Figure 4.4).

Behavioral performance by group

The top third of participants who had the highest accuracy scores across all three blocks were split into one group while the remaining participants were split into two groups based on the median difference between scores in the first and third blocks. Participants who showed a greater increase in accuracy from the start and end of the study were labeled 'improvers', while participants whose performance decreased or plateaued were called 'decliners' (Figure 4.5). Participants were surveyed for their musical background, and years spent playing their primary instrument was used as a measure of years of musical experience. Years of musical experience were significantly higher in consistent high performers compared to decliners ($t(16) = 2.59$, $p < 0.05$), while there was no difference in musical experience between the

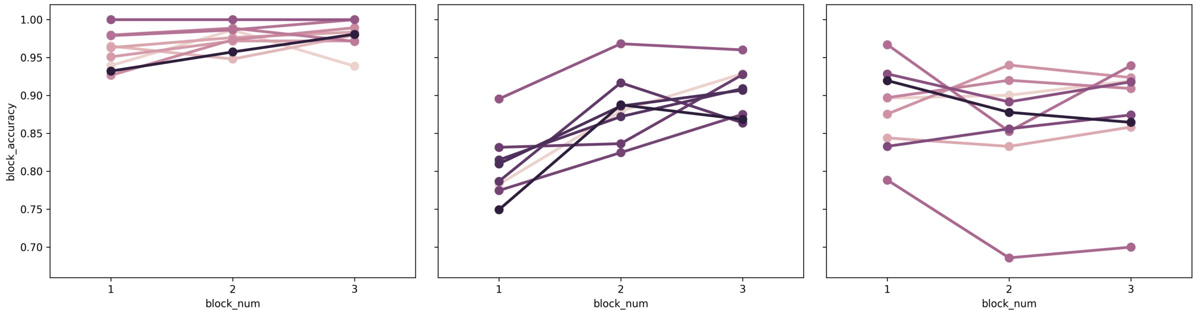


Figure 4.5: Participants were split into three groups based on the trajectory of their task performance. The first group consisted of consistent high performers (left), participants who showed greater improvement across blocks were labeled 'improvers' (middle), and participants whose performance plateaued or declined were called 'decliners' (right).

other groups. In comparison, improvers had lower Auditory N-Back scores than decliners (t(15) = 2.12, $p < 0.1$) and experts (t(15) = 1.93, $p < 0.1$) (Figure 4.6).

FFR results by group

A linear mixed-effects model was fitted to examine the effects of task modality (passive vs. active), task order (first vs. second), tone frequency (130, 200, 280 Hz), and performance trajectory group (expert, improver, decliner) on FFR amplitudes, with subject included as a random intercept. The model showed a significant main effect of tone frequency, F(2, 256) = 32.1, $p < .0001$. Estimated marginal means indicated that responses were largest for 200 Hz tones (M = 6.77 μV, SE = 0.49, 95% CI [5.53, 8.00]), followed by 280 Hz (M = 3.35 μV, SE = 0.49, 95% CI [2.12, 4.58]) and 130 Hz (M = 2.98 μV, SE = 0.49, 95% CI [1.75, 4.21]). Tukey-corrected posthoc comparisons confirmed that 200 Hz amplitudes were significantly larger than both 130 Hz (Δ = 3.79 μV, SE = 0.43, t(256) = 8.88, $p < .0001$) and 280 Hz (Δ = 3.42 μV, SE = 0.43, t(256) = 8.01, $p < .0001$), whereas 130 Hz and 280 Hz did not differ (Δ = -0.37 μV, SE = 0.43, t(256) = -0.87, p = .66) (Figure 4.7).

There was also a robust main effect of task modality, amplitudes were larger during passive listening (M = 5.65 μV, SE = 0.51, 95% CI [4.46, 6.84]) compared to active listening

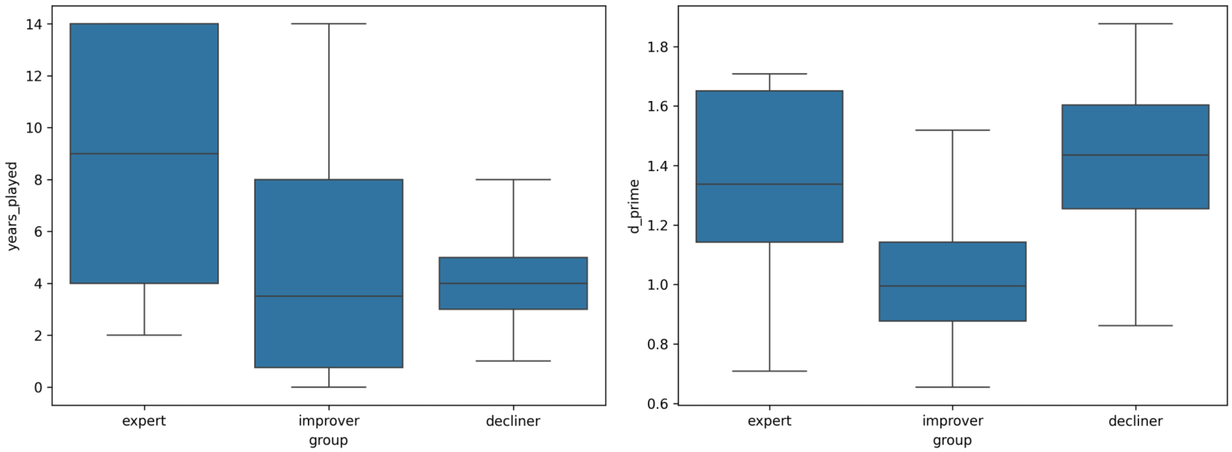


Figure 4.6: Duration of time playing a single instrument (left) and Auditory N-Back scores by performance trajectory group. Consistent high performers had more years of musical experience than decliners. Improvers had lower Auditory N-Back scores than consistent high performers and decliners.

(M = 2.55 μV, SE = 0.45, 95% CI [1.48, 3.62]), t(256) = -8.92, p < .0001. This effect was consistent across tones. For example, at 130 Hz, amplitudes were higher in the passive condition (M = 4.11 μV, SE = 0.65) than the active condition (M = 1.85 μV, SE = 0.49), Δ = -2.25 μV, SE = 0.60, t(256) = -3.73, p = .0002. Similar effects were observed for 200 Hz (passive: M = 8.90 μV, SE = 0.65; active: M = 4.63 μV, SE = 0.49; Δ = -4.27 μV, SE = 0.60, t(256) = -7.08, p < .0001) and 280 Hz (passive: M = 4.95 μV, SE = 0.65; active: M = 1.75 μV, SE = 0.49; Δ = -3.21 μV, SE = 0.60, t(256) = -5.32, p < .0001). In addition, task order interacted significantly with tone frequency. Specifically, amplitudes at 200 Hz and 280 Hz were reduced when the passive task occurred second (order = 2) compared to first, as reflected in significant task order × tone interactions (200 Hz: B = -3.76, SE = 1.48, t(256) = -2.53, p = .012; 280 Hz: B = -3.51, SE = 1.48, t(256) = -2.36, p = .019).

Regarding performance trajectory groups, although the main effects of group were not significant, there were trends toward reduced amplitudes for experts and improvers compared to decliners at 200 Hz (expert: B = -2.83, SE = 1.53, t(256) = -1.85, p = .066; improver: B = -2.70, SE = 1.53, t(256) = -1.76, p = .079). Critically, a significant three-way interaction emerged between task order, tone frequency, and group for improvers at 200 Hz (B = 5.52, SE = 2.13, t(256) = 2.58, p = .010). Posthoc comparisons revealed that improvers in the second passive-order condition exhibited the largest amplitudes (M = 11.24 μV, SE = 1.40, 95% CI [6.77, 15.71]), compared to improvers in the first passive-order condition (M = 6.16 μV, SE = 1.09), decliners in the second passive-order condition (M = 5.75 μV, SE = 0.99), and experts in the second passive-order condition (M = 3.74 μV, SE = 1.21).

No significant differences were observed among groups at 130 Hz or 280 Hz (all ps > .15). For example, at 130 Hz, decliners in order 1 (M = 3.57 μV, SE = 1.40) did not significantly differ from experts in order 2 (M = 0.72 μV, SE = 1.21) or improvers in order 2 (M = 5.64 μV, SE = 1.40), all ps .16. Similarly, at 280 Hz, amplitudes did not differ significantly across performance groups within either task order (all ps > .56).

In sum, the model revealed strong main effects of tone frequency and task modality, with larger amplitudes observed for 200 Hz tones and during passive listening. Performance trajectory groups showed trends at 200 Hz, and a significant interaction indicated that improvers, particularly when the passive condition was performed second, exhibited greater response amplitudes. No group effects were observed at 130 Hz or 280 Hz.

PLV results by group

An ANOVA was then computed over PLV values with tone frequency, task condition, and performance group as fixed factors. The observed PLV values were also computed by bootstrapping, which allows pairwise statistical tests to be conducted between groups (Figure 4.8). There was a main effect of performance trajectory group on PLVs (F(2, 50) = 10.30, p < 0.001). Consistent high performers had the lowest PLV values (M = 0.064, SD = 0.014), significantly lower than improvers (M = 0.078, SD = 0.025) (mean difference = 0.017, t(25) = 3.40, p = 0.003), and significantly lower than decliners (M = 0.083, SD = 0.036) (mean difference = 0.020, t(25) = 4.24, p < 0.001). Post-hoc comparisons indicate that this effect

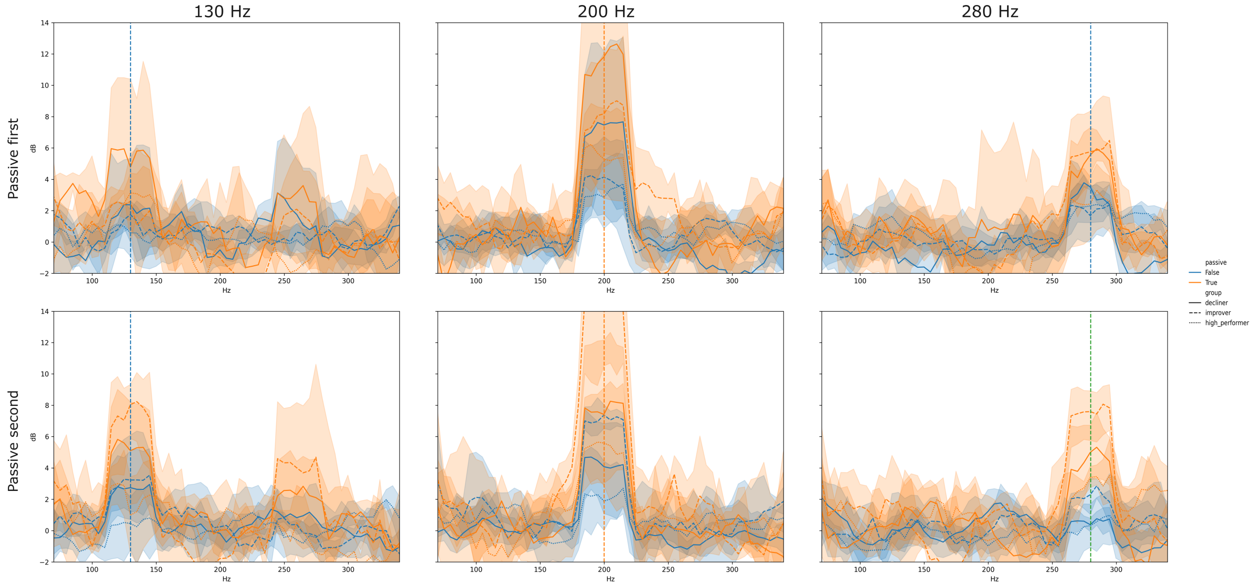


Figure 4.7: The same main effects of tone frequency and task condition are shown when categorizing participants by their task trajectory instead of their overall task accuracy. Grouping participants by task trajectory also revealed a strong interaction between task order and group, in addition to weaker interactions between task condition and group, task order, tone frequency, and group.

was mainly driven by the decrease in PLV values to the 200 Hz tone in experts (M = 0.073, SD = 0.020) compared to decliners (M = 0.106, SD = 0.047) (mean difference = 0.038, t(120) = 4.6, $p < 0.001$), and in experts compared to improvers (M = 0.100, SD = 0.030) (mean difference = 0.031, t(120) = 3.69, $p < 0.05$). There were significant interaction effects between task order and group (F(4, 102) = 2.18, p = 0.076), with significant pairwise contrasts between consistent high performers and improvers (mean different = 0.046, t(25) = 5.43, $p < 0.001$) and consistent high performers and decliners (mean different = 0.046, t(25) = 5.43, $p < 0.001$). There was also a main effect of tone frequency (F(2, 24) = 24.8, $p < 0.001$). PLVs to the 200 HZ tone (M = 0.093, SD = 0.037) (mean difference = 0.026, t(25) = 5.31, $p < 0.001$) were higher than to the 130 Hz (M = 0.069, SD = 0.021) and 280 Hz tones (M = 0.063, SD = 0.009) (mean difference - 0.033, t(25) = 6.66, $p < 0.001$). PLVs between the 130 Hz and 280 Hz tones were not significantly different (mean difference = 0.007, t(25) = 1.35, $p > 0.1$). There was also a significant interaction between task order

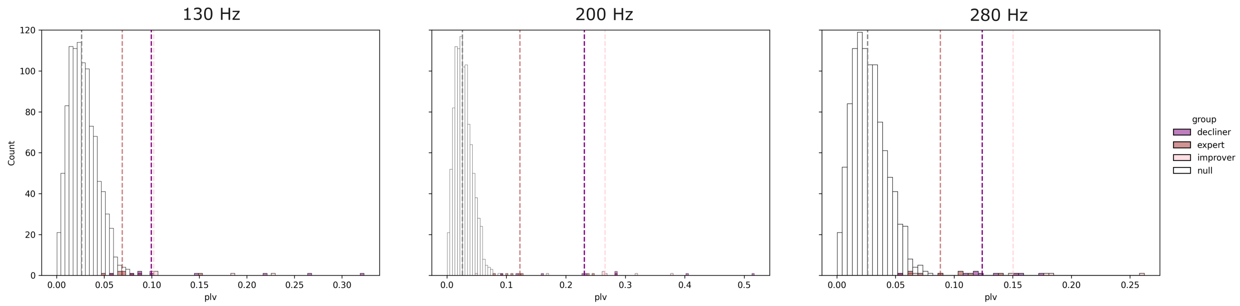


Figure 4.8: Observed PLV values for all groups were above null. PLVs for consistent high performers were significantly lower than for both improvers and decliners for the 200 Hz and 280 Hz tones.

and performance trajectory, although this interaction did not reflect any interpretable or theoretically meaningful pattern and is not discussed further.

4.4 Discussion

This study extends the findings from previous experiments by introducing a passive listening condition to examine the joint effects of task performance, effort, and attention on auditory encoding. While the first experiment found increased FFR amplitudes in high performers during an intermodal attention task, the second found decreased amplitudes in a simpler task. The present study offers a more nuanced perspective by comparing active and passive conditions within the same participants to see if low effort and attention listening conditions increase FFR amplitudes as predicted by dynamic tuning accounts, or decrease them as predicted by gain modulation. Critically, both prior studies found no significant effects of task demands or interactions with performance, suggesting that attentional instructions alone may be insufficient to modulate early auditory encoding. These findings highlight the need to distinguish between effort-driven gain modulation and experience-driven encoding efficiency. By including both active and passive conditions, this study evaluates how attentional demands and task difficulty interact with individual differences in expertise and performance to shape neural responses, addressing a notable gap in the FFR literature where passive paradigms have been used almost exclusively to assess expertise effects.

The present study replicates the decrease in FFR amplitudes in high task performers found in Experiment 2. As with Experiment 2, behavioral results indicated that the current task was relatively easy, with high overall accuracy levels including multiple participants at ceiling. This suggests that cognitive effort was minimal for many participants. According to the FUEL framework (Pichora-Fuller et al., 2016), easier tasks require less allocation of cognitive resources, particularly in individuals with greater expertise or familiarity. This is consistent with measures of listening effort in musicians and non-musicians, which report reduced pupil dilation and neural effort when in musicians when listening to music (Endestad et al., 2020). The decrease in FFR amplitude among high performers supports the predictions from dynamic tuning accounts, which suggest that reduced effort and increased response selectivity, rather than degraded processing, can yield attenuated neural responses. Dynamic tuning models (e.g., Fritz et al., 2007; Schreiner & Polley, 2014) propose that attention does not simply increase the overall magnitude of neural responses (as in gain models), but instead enhances selectivity by sharpening receptive fields and suppressing responses to irrelevant or redundant information. As such, high performers may be able to maintain accuracy using more selective and efficient encoding strategies, thereby reducing the magnitude of neural responses while preserving encoding fidelity.

The results from this study also show a decrease in FFR amplitudes during active listening relative to passive listening. The decrease in FFR amplitudes during active listening is a surprising result given that past studies generally find an increase in FFR amplitude or consistency under active listening. However, most of these studies were conducted using very short stimuli durations (Galbraith & Kane, 1993; Hoorman et al., 1994, 17.5 ms and 30 ms long stimuli), and Experiment 1 has shown that amplitude effects are difficult to measure using short stimuli. A more recent study by Price and Bidelman (2021) also found an increase in FFR amplitudes under active listening compared to passive listening, using continuous speech stimuli with varying amounts of noise added to manipulate task difficulty. The difficulty of the task may have increased effortful listening in participants, explaining the increase in FFR amplitude that they observed. Moreover, they used speech stimuli that are highly behaviorally salient and likely changed how participants attended to the stimuli relative to pure tones.

The decrease in FFR amplitudes in this study is consistent with predictions from dynamic tuning models about the mechanisms of auditory attention under passive versus active listening conditions. Dynamic tuning models (Fritz et al., 2007; Schreiner & Polley, 2014) propose that attention acts by refining receptive fields and enhancing stimulus selectivity, rather than increasing neural gain globally. This predicts that active listening, particularly for high performers in easier tasks, leads to sharper tuning and suppression of irrelevant features, resulting in smaller, more focused neural responses. In contrast, passive listening, which lacks top-down attentional selection, may elicit broader, less selective neural activity and higher FFR amplitudes. This interpretation aligns with the current findings, suggesting that dynamic tuning, not gain modulation, dominates under conditions of low cognitive demand.

While the observed decreases in FFR during passive listening relative to active listening are consistent with dynamic tuning models, it is inconsistent with some results from past FFR studies. Past FFR studies on attended versus unattended acoustic stimuli typically report null results. However, these null results may be attributed to the short stimuli duration used in these experiments. For example, Hoorman and colleauges (1994) had participants listen to pairs of pure tone stimuli, the first tone was always a 340 Hz tone with 10 cycles (30 ms). On rare occasions, the second tone would be a target tone with a higher pitch than the original stimulus. During active blocks, participants were asked to count the number of target tones, while in passive blocks, they were not required to respond to the tones. Hoorman and colleagues reported no significant amplitude effects and only a slight decrease in FFR latency during the active task. Hoorman and colleagues (2000) replicated this paradigm, adding additional intermodal and dichotic listening tasks, and also found no amplitude effects on attended compared to unattended tones. Galbraith and Kane (1993) report similar null results in a task with sequences of 17.5 ms, 230 Hz tones under an active task with complex auditory targets and a close-eyed passive listening task. As established in Experiment 1 of this dissertation, short stimulus durations reduce temporal resolution and are generally insufficient for accurately measuring the amplitude of the resulting stable-state response. An exception to this trend in null results is a study by Price and Bidelman (2021) in which participants listened to either a clean stream of synthesized vowels or synthesized vowels embedded in an 8-talker babble. They reported increased FFR amplitudes in participants during blocks in which they were asked to respond to infrequent vowel tokens, compared to FFR amplitudes during passive listening tasks accompanied by a silent movie. While a direct replication and comparison between the paradigms used in this study is warranted to draw any reliable conclusions, the opposite effect observed by Price and Bidelman may be due to the use of speech stimuli, which are known to elicit distinct cortical areas from non-speech sounds (Binder et al., 2000; Jaramillo et al., 2001). Moreover, their use of stimuli containing only 15 cycles of the F0 of the stimulus signal may limit the reliability of their power estimates.

The significantly higher FFR amplitudes observed during the passive task also diverge from past FFR literature on expertise. Studies on FFRs in experts, such as musicians and bilinguals, typically report an increase in FFR amplitudes or consistency in experts under passive listening conditions. Again, one possible source of this difference is the use of speech stimuli in this literature (e.g., Wong et al., 2007; Tierney et al., 2013; Tierney et al., 2015; Song et al., 2008; Russo et al., 2005). However, such a marked difference in the neural mechanisms engaged between speech and non-speech stimuli is not predicted by any current framework of auditory attention that is concerned with how neural firing rates encode acoustic features. These differences in results encourage a replication of the current experiment with the addition of a passive listening task using the /da/ stimuli commonly used in this body of literature (Tierney et al., 2013; 2015; Russo et al., 2005; Krizman et al., 2012; Kraus et al., 2017).

The results presented in this study, in conjunction with those reported in the previous two studies, demonstrate a flexible use of gain modulation and feature-selective tuning in audition depending on effort and task demands. Under difficult active listening conditions, increased effort appears to elicit gain-modulated increases in FFR amplitude, supporting improved encoding through heightened neural responsiveness. Under easier active listening conditions, however, experienced listeners may engage dynamic tuning strategies, reducing the recruitment of neurons while preserving encoding precision. Passive listening, on the other hand, does not require the suppression of any part of the stimulus stream, and no top-down influence is exerted on early auditory processes. As a result, neural activity can be freely patterned by oncoming stimuli and bottom-up auditory activity to unfold without suppression, producing greater neural responses but potentially with less selectivity. The lack of any significant difference in PLV between conditions and between task performance groups suggests that the consistency of phase looking is quite stable between conditions and task performance; rather, it is the degree of neural recruitment involved that changes. Taken together, the three experiments presented so far in this dissertation suggest that active listening and experience can reduce neural responses through dynamic tuning, but that effort and gain modulation can increase responses again through effort under difficult task conditions.

This study also included an analysis in which participants were grouped not based on their overall performance, but based on the trajectory of their performance throughout the three blocks of the active task. This allowed for finer-grained distinctions between the effects of expertise and effort. Participants who were consistently high performing or were at ceiling through the task and presumably found the task easier and expended less effort. Consistent with FUEL and dynamic tuning, these participants exhibited significantly lower FFR amplitudes to all tones than improvers and decliners. Improvers displayed higher FFR amplitudes than consistently high performers, particularly in the passive condition, possibly reflecting greater attentional recruitment and the engagement of gain modulation mechanisms, even during passive blocks. Interestingly, improvers who received the passive task first show a decrease in FFRs relative to decliners during both passive and active tasks. This decreased FFR during the active task may reflect their relative increase in encoding efficiency and selective activity compared to decliners. There are various possible explanations for the reduced FFRs in improvers relative to decliners during the passive task. One possible explanation is that improvers are highly conscientious study participants who listened actively during the passive listening task, thereby engaging selective, dynamic tuning mechanisms. Moreover, the difference between FFR amplitudes between improvers and decliners flips in participants who received the passive task second. It is possible that receiving the active task first primed participants to attend to the stimulus tones in a certain way, which carried into the passive task. In other words, improvers may be applying an effortful attention, through gain-modulated activity, to the active task that they transfer into the passive listening task. Splitting participants into groups based on their performance trajectories corroborates with accounts of auditory attention, which proposes that effort engages gain modulation and increases in neural sensitivity accentuate auditory encoding during difficult tasks. These findings lend further support to the view that gain modulation and dynamic tuning operate as complementary systems, recruited flexibly depending on task demands and individual effort levels.

Lastly, in the passive listening condition, high performers exhibited increased FFR amplitudes relative to their active task responses, but the interactions between performance group and condition suggest that this increase was not as pronounced as in other participant groups. As with the previous two experiments, task performance was significantly correlated with years of musical experience. Musical experience was measured as the greatest number of years playing a single instrument. This pattern suggests that while passive listening broadly enhances FFR amplitude due to the absence of attentional suppression, individuals with greater expertise may still engage more selective neural encoding even when not actively attending to the stimuli. In other words, the auditory systems of high performers appear to be shaped by long-term experience in a way that promotes efficient and selective processing, which generalizes beyond specific task contexts. This reduced amplitude increase relative to lower performers may reflect a form of experience-driven neural tuning in which irrelevant features are automatically suppressed, even in the absence of an explicit task. These findings support the notion that auditory expertise leads to stable, context-independent changes in sensory encoding, consistent with dynamic tuning accounts of experience-dependent plasticity (Fritz et al., 2007; Schreiner & Polley, 2014).

Together, these results offer converging evidence that high performers on an easy auditory task likely engaged reduced cognitive effort, enabling them to rely on more efficient and automatic processing strategies. The observed decrease in FFR amplitude among these individuals is consistent with dynamic tuning models, which propose that expertise enhances neural selectivity by suppressing irrelevant activity rather than boosting overall response strength. Furthermore, passive listening conditions lack explicit task demand and appear to elicit broadly increased FFR amplitudes across participants, reflecting a more diffuse pattern of neural engagement due to reduced selective attention. Even in the absence of task goals, high performers exhibited smaller amplitude increases than other groups, suggesting that long-term auditory experience results in stable, context-independent tuning that favors selective, efficient encoding. These findings underscore the importance of distinguishing between cognitive effort and encoding efficiency in auditory neuroscience, highlighting how both immediate task demands and long-term expertise shape the dynamics of early sensory processing through gain modulation and dynamic tuning.

CHAPTER 5

EXPERIMENT 4: TESTING THE BOUNDARY CONDITIONS OF ATTENTION MECHANISMS IN EARLY AUDITORY ENCODING

5.1 Introduction

Across the previous experiments, results suggest that high-performing individuals on an easy auditory task exhibited reduced FFR amplitudes, suggesting that they required less cognitive effort and were able to engage more efficient, automatic processing strategies. These findings align with dynamic tuning models, which propose that expertise enhances auditory encoding not by amplifying neural responses but by increasing neural selectivity and suppressing irrelevant activity. In contrast, passive listening conditions, which lack explicit task demands, elicited broadly elevated FFR amplitudes across participants, indicating a broader, non-selective pattern of neural engagement due to the absence of attentional filtering. Even under passive conditions, high performers showed smaller amplitude increases than other groups, suggesting that long-term auditory experience produces stable, context-independent changes in neural tuning that increase encoding efficiency. Together, these results highlight the operation of distinct neurobiological mechanisms, including dynamic tuning and gain modulation, which are recruited flexibly depending on task demands, listening effort, and individual differences in expertise.

A key question raised by this model is under what precise contexts these different mechanisms are activated. Prior findings suggest that active listening under easy task conditions engages dynamic tuning processes, which sharpen neural selectivity and reduce overall response magnitude, leading to decreased FFR amplitudes. In contrast, passive listening, which lacks explicit task demands or top-down attentional filtering, elicits the highest FFR amplitudes observed, consistent with diffuse, non-selective neural activation. These contrasting outcomes raise the question of whether the brain engages these mechanisms in a discrete, modular fashion, by switching between modes depending on context, or whether their recruitment varies along a continuum shaped by gradations of effort and attention. Boundary conditions, such as active tasks that are so easy they require minimal cognitive effort that their attentional demands approach those of passive listening, may help adjudicate this issue. Addressing cases like these may help refine models of the conditions under which neural gain or tuning mechanisms dominate early auditory processing.

To probe the boundary between passive and active listening states, the present experiment introduces an extremely easy variant of the active oddball task used in the previous studies. In this condition, tones are presented in a regular, repeating sequence ("AABBCCAAB…"), dramatically reducing task complexity. Behavioral results confirm that participants performed at ceiling, suggesting minimal cognitive effort was required to complete the task. This design enables a critical test of whether the distinction between passive and active listening reflects a discrete, modular shift in processing modes or a more continuous gradient shaped by attentional demands. If dynamic tuning can be engaged more incrementally, then this almost effortless active task may elicit higher FFR amplitudes that approach those observed during passive listening, reflecting reduced neural selectivity. In contrast, if active listening engages mechanisms of selective attention regardless of task difficulty, then even highly predictable active tasks should exhibit reduced FFR amplitudes, consistent with dynamic tuning, gain modulation, and predictive coding models, all of which predict decreased neural responses under predictable or low-effort, active listening conditions. Thus, only if FFR amplitudes increase under this extremely easy active task, similar to passive listening, can a distinction be made favoring one model over others. Increased FFR amplitudes to the predictable tone sequences would suggest that passive, non-selective neural engagement may also be recruited under minimal cognitive load during active states.

As with Experiment 2, ERPs will be simultaneously recorded with FFRs. In contrast to FFRs, which measure rapid, early, phase-locked responses along different centers of the auditory pathway, ERPs generally reflect late, low-frequency, and large-amplitude responses in the cortex. Comparing the two responses can help localize where in the auditory hierarchy task demands or experience exerts the greatest influence. ERP components such as the N1, P2, N2, and P3b are associated with stimulus detection, categorization, attention allocation, and decision-making, and primarily reflect cortical processing. As an exploratory study, this experiment does not make specific predictions about individual ERP components and their correlations with changes in FFR amplitude and consistency. Instead, this study focuses simply on the presence or absence of ERP and FFR effects and the types of top-down changes in auditory activity they reflect.

In addition to replicating the main effect of task performance on FFR amplitudes during an active listening task for the second time, this study underscores the complexity involved in delineating the boundaries between different neurobiological mechanisms of auditory encoding. While the framework outlined in this dissertation offers a valuable account of how gain modulation and dynamic tuning may be differentially engaged based on task performance, effort, and task demands, the current findings also highlight its limitations. In particular, the observation of elevated FFR amplitudes under passive listening conditions, especially among non-experts, raises questions about whether broader or non-selective mechanisms engaged typically during passive listening may also be recruited under very low effort, active listening conditions. Although this possibility aligns with predictions from dynamic tuning models, it remains underdeveloped in current literature. In short, this study is included in this dissertation to emphasize the need for continued replication and systematic testing of the boundaries of the current model. A careful, incremental approach will ensure that this framework remains both flexible and theoretically coherent as new data emerges.

5.2 Materials and methods

32 participants (17 women, 12 men, 3 non-binary) were recruited from the University of Chicago community for this experiment. None of the participants had taken part in the previous experiments. All participants were confirmed to have normal speech and hearing, as assessed using a Welch Allyn otoscope. Data from two participants were excluded due to excessive artifact rejection, specifically, fewer than 80% of their FFR trials were retained after applying a 35 mV peak-to-peak rejection threshold to the PREP-preprocessed data (Bigdely-Shamlo et al., 2015). Informed consent was obtained from all participants in accordance with protocols approved by the University of Chicago Institutional Review Board. Participants received either course credit or $20 per hour in compensation.

Stimuli and task

In two of the four blocks the tones were presented in a repeating AABBCC pattern, and in the remaining two the tones were presented randomly. The stimuli used in this task were three pure sine-wave tones at 110, 150, 210 Hz. Tones onsets and offsets were smoothly ramped with a 5 ms Hanning window. The tones were presented at 65–70 dB SPL over insert earphones (Etymotic ER3C), the tones were 300 ms in duration with a jittered interstimulus interval from 200-250 ms. The length of sequences varied uniformly between 42 and 50 tones. One of the three tones was randomly selected to be the target tone at the start of each trial. Before each trial, participants were allowed to listen to the target tone as many times as they wished. Once the trial started, participants were asked to count the number of times the target tone played and report their answer at the end of the trial. Tone sequences were pseudo-randomized such that the first tone was never a target tone, and no tone would play more than three consecutive times.

As with Experiments 1 and 2, participants were given one ‘point’ if they answered correctly, or if their guess came within 2 tones of the actual target count. Each block consisted of 18 trials and participants completed 4 blocks each, with block conditions pseudorandomized such that odd-numbered participants completed the blocks with randomized tones in the first and third blocks, while even-numbered participants completed the blocks with randomized tones in the second and fourth blocks. The entire task lasted approximately 1 hour with 700-900 sweeps per tone per condition. Participants were told they could earn up to an additional $10 in bonus compensation based on their performance, though all participants ultimately received the bonus. Accuracy for each trial was calculated by subtracting the ratio of the counting error from 1, where the error was defined as the absolute difference between the reported and actual number of targets. Each participant's overall performance was then quantified as the average accuracy across the first 18 trials of every block. Based on the median accuracy across all participants, participants were classified into two groups: those scoring below the median were designated as low performers, while those above the median were labeled high performers.

Electrophysiological data acquisition and preprocessing

Prior to EEG cap placement, participants were told what to expect from the task and EEG procedure. The circumference of each participant's head was measured to fit the actiCAP slim EEG 64-channel net (actiCAP, Brain Products GmbH, Germany). Participants were seated in a bright, sound-attenuated, and electromagnetically shielded room and asked to attend to the tones while minimizing eye blinks and other movement artifacts. After the experiment, the exact electrode positions were recorded using CapTrak (Brain Products GmbH, Germany). EEG data were collected with a sampling rate of 5 kHz using an actiCHamp Plus amplifier (Brain Products GmbH, Germany). Two electrodes (originally AF7 and AF8 in the actiCAP layout) were affixed under the outer canthi of the left and right eyes to monitor eye movements (i.e., to be used as EOG channels).

The EEG data were preprocessed using custom Python scripts and the MNE-Python library version 1.5.0 (Gramfort et al., 2013). The eye electrodes (AF7 and AF8) were re-referenced to Fp1 and Fp2, respectively (i.e. below the eye referenced to above the contralateral eye), to create bipolar EOG channels. The automated PREP pipeline was used to notch filter the data at 60 Hz, identify and exclude bad channels using threshold-based artifact rejection, and re-reference the EEG data to the average of all electrodes. Detailed participant-level preprocessing reports containing details such as trial counts, removed Independent Components, interpolated electrodes, and data visualizations for quality check are available with our dataset on OpenNeuro. As in Experiments 1 and 2, a timing test was run before the experiment to measure the lag and jitter between event tags and stimulus onset. The computed jitter was 0.67, and the computed lag of 27.2 ms was used to correct event tags prior to further analysis. The two experiments were conducted several months apart.

FFR analysis

Preprocessed data from all channels except Cz were dropped for FFR analysis. Since tasks in this experiment used only stimuli of a single polarity, the common average reference applied during PREP preprocessing was maintained for the FFR analysis. Using a common average reference mitigates the likelihood of biasing responses towards peripheral sources, including the cochlear microphonic and auditory nerve, and overestimating the absolute amplitude of the signal (Bidelman et al., 2015; Coffey et al., 2019). The data were epoched from -200 ms to 400 ms relative to stimulus onset. The epoched data were then time-shifted by -27.2 ms based on the lag in stimulus presentation calculated as the difference between the timestamp of each event tag and the onset of the stimulus signal as recorded by the StimTrak. The epochs were then bandpassed from 50 to 300 Hz using an FIR filter with a lower and higher transition bandwidth of 12.5 Hz and 75 Hz and a filter length of 6.6 times the lower transition bandwidth. To compute the power at each stimulus frequency for each participant, trials for each tone and target condition were first averaged in the time domain.

Next, the power spectral densities of the 200 ms baseline period and the first 200 ms following stimulus presentation for the averaged signal were calculated. Finally, the logarithm of the power ratio between the baseline and stimulus periods was multiplied by 10 to yield power in decibels (dB).

PLV analysis

The consistency of phase locking in the neural response was assessed using the phase-locking value (PLV; Lachaux et al., 1999; Zhu et al., 2013). FFR epochs were cropped to include only the segment following stimulus onset. For each participant and condition, PLV was computed via bootstrapping: 400 trials were randomly sampled, and the fast Fourier transform (FFT) was applied to each. Phase values at each stimulus frequency were extracted by isolating the angle of the complex-valued FFT output. The PLV for each draw was then calculated as the magnitude of the average of the complex exponentials of the phase angles across the 400 trials. This process was repeated 1000 times to obtain a stable estimate of PLV per stimulus per participant. A null distribution for statistical testing was generated through phase-scrambling, following Zhu et al. (2013), by computing PLVs from 1000 sets of randomly generated phases, repeated 1000 times. Statistical significance was assessed by comparing observed PLVs to the null distribution and by conducting pairwise comparisons between high- and low-performing participants, as well as across conditions.

ERP analysis

EEG data used in this analysis were preprocessed following the procedures described in Experiments 1 and 2. PREP-processed data were bandpass filtered from 0.1 to 40 Hz using a finite impulse response (FIR) filter with transition bandwidths of 0.1 Hz (low) and 10 Hz (high), and a filter length equal to 6.6 times the lower transition bandwidth. The continuous data were then segmented into epochs extending from –200 ms to 500 ms relative to stimulus onset and downsampled to 1 kHz. To remove ocular and muscle artifacts, independent component analysis (ICA) was applied to the preprocessed EEG signal, decomposing the data into 15 independent components (ICs). Components reflecting electrooculographic (EOG) artifacts were identified by correlating each IC time course with the EOG channels and z-scoring the resulting correlation coefficients. Any component with a z-score exceeding ±1.96 was excluded from the EEG signal. Trials were baseline corrected using the –200 to 0 ms pre-stimulus interval and automatically rejected based on peak-to-peak amplitude thresholds using the autoreject package (Mainak et al., 2017). Finally, the remaining epochs were averaged by condition and target tone to compute the time-domain event-related potentials (ERPs).

Amplitudes of ERP components were quantified by computing the mean signal amplitude within predefined time windows. These windows were determined through visual inspection of the averaged waveforms and based on previously published component timing conventions (Key et al., 2005). Specifically, the N1 component was measured from 100–150 ms, the P2 from 150–230 ms (Coch et al., 2005), the N2 from 250–360 ms (Folstein et al., 2007; Nieuwenhuis et al., 2003), and the P3b from 360–420 ms (Lange, 2009; Dien et al., 2010). Latencies were calculated by identifying the time point corresponding to the peak amplitude (minimum or maximum) within each window. For the smaller peak or inflection observed in the P3b window, latency was identified as the zero-crossing of the first derivative of the discrete time series. In Experiment 2, the P3b time window was adjusted to 350–420 ms to accommodate an earlier peak onset observed in the data.

5.3 Results

Behavioral results

An ANOVA was run on task scores with target identity and predictability as factors. As with the previous experiment, there was a main effect of target identity ($F(2, 58) = 9.01$, $p < 0.001$). Participants performed significantly worse when the middle tone was the target tone ($M = 89.4\%$, $SD = 12.4\%$) compared to the other tones (110 Hz $M = 93.6\%$, $SD = 6.4\%$; 210 Hz $M = 94.9\%$ $SD = 4.4\%$). The mean difference between performance to the 110 Hz tone compared to the 150 Hz tone was 4.2% $p < 0.01$, while the mean difference between performance to the 150 Hz tone compared to the 210 Hz tone was 5.5% $p < 0.001$ (Figure 5.1). An Ordinary Least Squares regression was run between years of musical experience and performance accuracy. Musical experience was measured as the duration of time, in years, spent playing a single instrument. The model was weakly statistically significant $F(1, 29) = 4.9$, $p < 0.05$, $R^2 = 2.1\%$ with years of musical experience significantly predicting scores $b = 0.3\%$, $p < 0.05$.

FFR results

A linear mixed-effects model was fitted to examine the effects of tone frequency (110, 150, 210 Hz), predictability (predictable vs. unpredictable), and task performance (high vs. low accuracy) on FFR amplitude, with subject included as a random intercept. The model revealed a significant main effect of tone frequency, such that FFR amplitudes were largest for 210 Hz tones, followed by 150 Hz, and smallest for 110 Hz tones. Posthoc Tukey-corrected comparisons confirmed that 210 Hz tones ($M = 5.09\ \mu V$, 95% CI [3.86, 6.32]) produced significantly greater responses than both 110 Hz ($M = 2.78\ \mu V$, 95% CI [1.54, 4.01]; $\Delta = -2.31\ \mu V$, $SE = 0.28$, $t(500) = -8.42$, $p < .0001$) and 150 Hz tones ($M = 3.52\ \mu V$, 95% CI [2.28, 4.75]; $\Delta = -1.57\ \mu V$, $SE = 0.28$, $t(500) = -5.72$, $p < .0001$). Additionally, 150 Hz tones

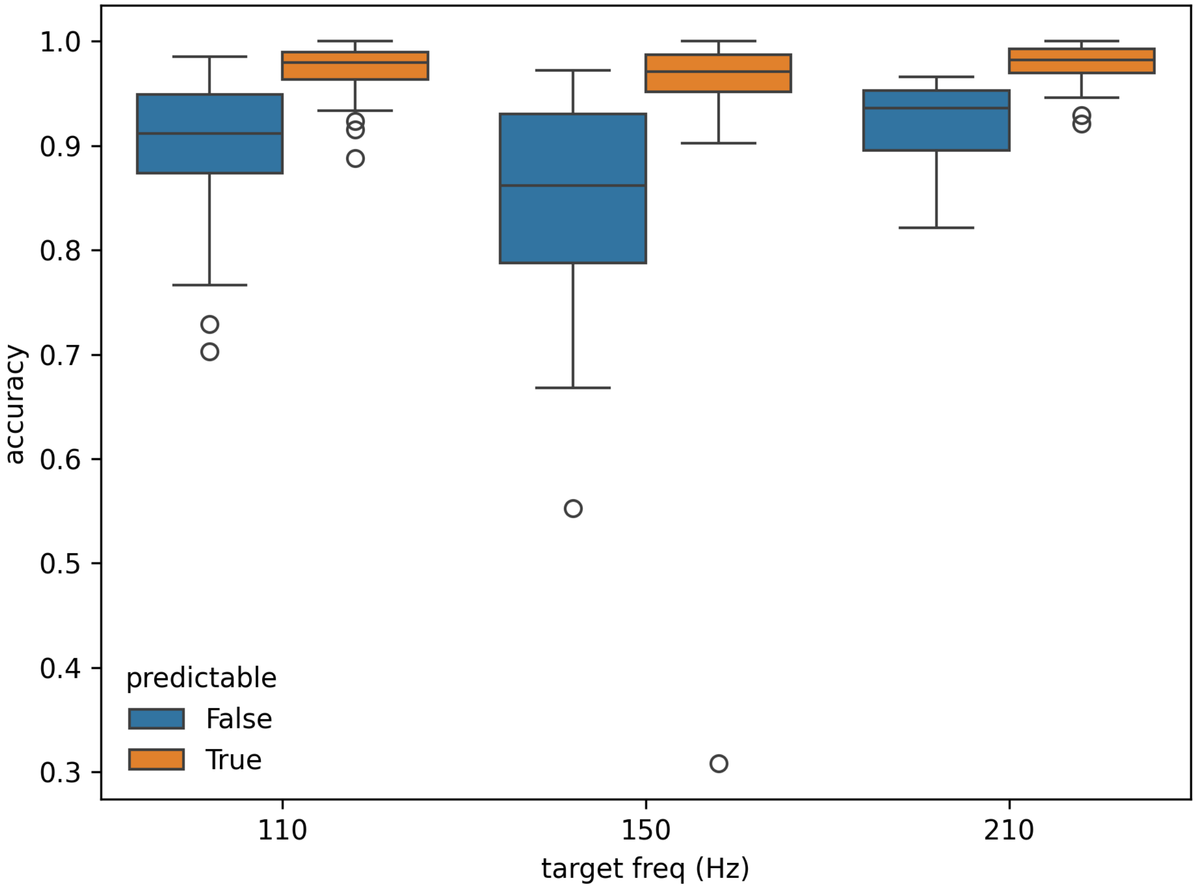


Figure 5.1: Behavioral results from both conditions. Participants performed significantly better in the blocks in which tones were presented in a regular pattern, with most participants at ceiling. As with Experiments 2 and 3, participants found the middle target tone the most difficult during the active task.

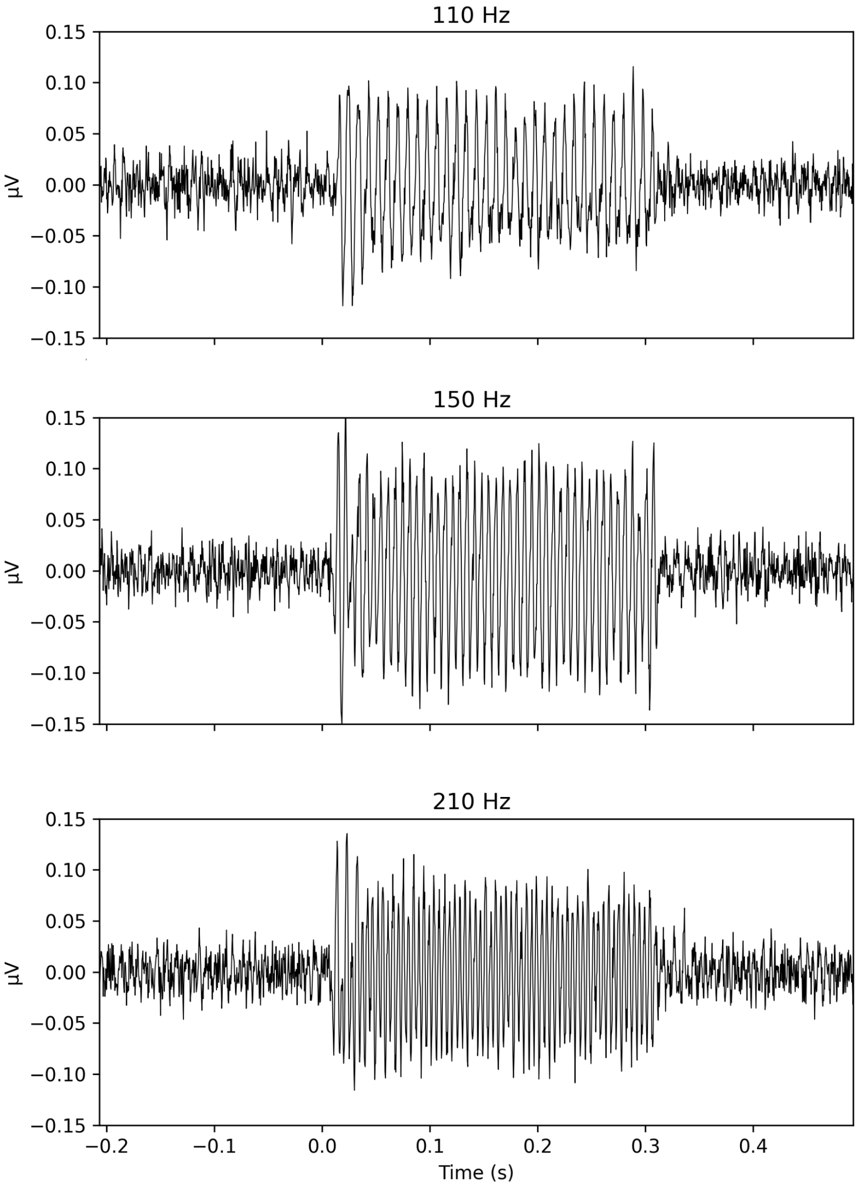
elicited significantly larger amplitudes than 110 Hz tones (Δ = -0.74 μV, SE = 0.28, t(500) = -2.70, p = .020). Models and posthoc comparisons were computed using the lme4 (Bates et al., 2015) and emmeans (Lenth, 2025) packages on R respectively (Figures 5.2 and 5.3).

A significant interaction between tone frequency and task performance also emerged. Follow-up contrasts indicated that for 210 Hz tones, low-performing participants exhibited significantly larger amplitudes (M = 6.55 μV, 95% CI [4.55, 8.56]) than high performers (M = 3.63 μV, 95% CI [1.75, 5.50]; Δ = -2.92 μV, SE = 0.98, t(34.8) = -2.97, p = .0054). No performance differences were observed at 110 Hz or 150 Hz (both ps > .30). Neither predictability nor its interactions reached significance.

Because target frequency was nested within tone conditions, a second model was run including target frequency as a factor. This model revealed no significant main effects or interactions of target frequency, predictability, or task performance (all ps > .10). Thus, while the first model demonstrated robust effects of tone frequency and a performance-related amplitude increase for 210 Hz tones, these effects did not extend to differences across specific target frequencies within tones.

PLV results

To determine whether the observed differences in FFR amplitude were attributable to reduced phase consistency, PLV values were computed separately for each tone for both high- and low-performing participants. PLVs were estimated using a bootstrapping procedure that involved randomly sampling trials, while null distributions were generated by randomly sampling phase values. Because the majority of PLV values were not normally distributed, statistical significance was assessed by comparing the median observed PLV to the 95th percentile of the null distribution (0.055) and the 90th percentile (0.048). Shapiro-Wilk tests confirmed the non-normality of the observed PLV distributions; therefore, Mann–Whitney U tests were used to compare PLV values between performance groups (Figure 5.4).


110 Hz

150 Hz

210 Hz

Figure 5.2: Grand average FFRs in the time domain for the tones in Experiment 2. Both the transient and sustained components of the early auditory response can be seen in the signals.

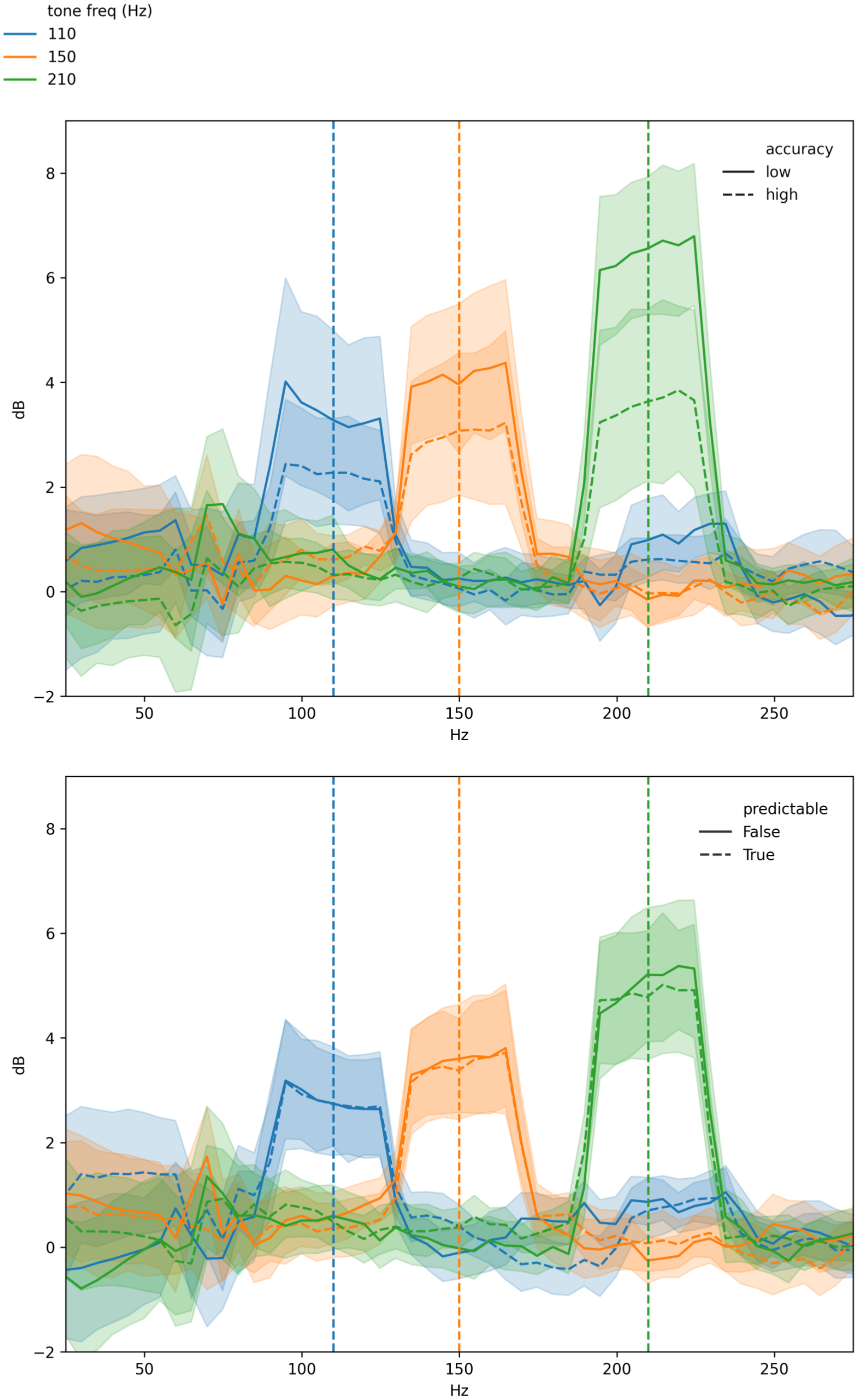


Figure 5.3: FFR amplitudes to each tone. Lines show the mean power in dB in the evoked response to each tone relative to the baseline period; the shaded regions represent the 95% CI of the mean. There was no effect of predictability on FFR amplitudes (middle). Experiment 2 reproduces the effect of task performance on FFR amplitudes. High task performers had much lower magnitude spectra than low task performers (bottom). There were no significant interaction effects.

For the 110 Hz tone, both low performers (median = 0.051, IQR = 0.008, $p < 0.1$) and high performers (median = 0.056, IQR = 0.010, $p < 0.05$) showed significantly greater PLVs than the null distribution. However, PLVs did not differ significantly between groups (z = 132, $p > 0.1$). Similarly, for the 150 Hz tone, PLV values were significantly greater than null for both low performers (median = 0.054, IQR = 0.010, $p < 0.1$) and high performers (median = 0.053, IQR = 0.0.008; $p < 0.1$), with no significant group difference (z = 101, $p > 0.1$). Finally, for the 210 Hz tone, PLVs again exceeded the null distribution for both low (median = 0.058, IQR = 0.009, $p < 0.05$) and high performers (median = 0.053, IQR = 0.013; $p < 0.1$), with no difference observed between groups (z = 90, $p > 0.1$).

Similar patterns were observed when comparing PLV values to predictable and random tone sequences. Both PLVs for FFRs to both predictable (median = 0.049, IQR = 0.009, $p < 0.1$) and random (median = 0.053, IQR = 0.011, $p < 0.1$) and were significantly greater than the null distribution. PLVs did not differ significantly between groups (z = 333, $p > 0.1$). Similarly, for the 150 Hz tone, PLV values were significantly greater than null for both predictable (median = 0.052, IQR = 0.010, $p < 0.1$) and random tone sequences (median = 0.055, IQR = 0.011; $p < 0.05$), with no difference observed between groups (z = 408, $p > 0.1$). Finally, for the 210 Hz tone, PLVs again exceeded the null distribution for both predictable (median = 0.056, IQR = 0.017, $p < 0.05$) and random tone sequences (median = 0.054, IQR = 0.010; $p < 0.1$), with no difference observed between groups (z = 433, $p > 0.1$).

ERP results

There was a significant effect of task performance on the mean amplitudes of nearly all of the ERP components. The main effect of accuracy on the P2 was $F(1, 28) = 4.29$, $p < 0.05$, the main effect of accuracy on the N2 was $F(1, 28) = 17.88$, $p < 0.001$, and the main effect of accuracy on the P3b was $F(1, 28) = 5.94$, $p < 0.05$. Finally, there was also a small

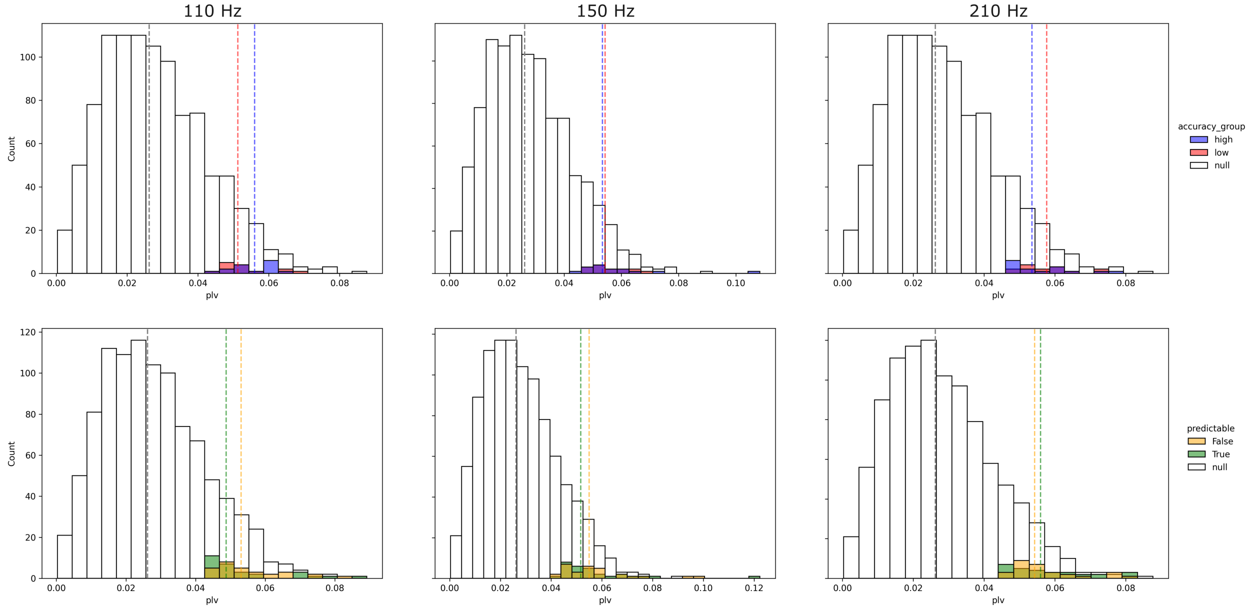


Figure 5.4: PLVs were slightly higher than chance for both high and low task performers, with no difference in PLV values between the two groups (top). PLVs were also slightly higher than chance for random and predictable tone sequences, with no difference in PLV values between the two conditions (bottom).

but significant interaction between all four variables of tone frequency, target frequency, predictability, and accuracy on mean amplitude in the P3b component (F(4, 118) = 2.09, $p < 0.1$). Unlike Experiment 2, however, there was no accuracy effect on the N1, but there was a significant accuracy effect on the late P3b component (Figure 5.5).

There was also a main effect of predictability in the two negative-going components N1 (F(1, 28) = 6.15, $p < 0.05$) and N2 (F(1, 28) = 46.71, $p < 0.001$). There was also a small but significant interaction between predictability and tone frequency on the mean amplitude of the N2 (F(2, 58) = 4.83, $p < 0.01$) and the P3b (F(2, 58) = 2.63, $p < 0.1$), suggesting that the effect of predictability depends on tone frequency.

As with Experiment 2, there was also a significant interaction between tone frequency and target frequency on late ERP components (Figure 5.5). The interaction between tone and target suggests that changes in cortical responses occur when a tone is selected as a target tone. There was an interaction effect between tone and target frequency on the mean amplitude of the N1 component measured between 230-360 msec ($F(4, 118) = 5.10$, $p < 0.001$), and on the P3b component measured between 360-420 msec ($F(4, 118) = 11.91$, $p < 0.001$). There was also a main effect of tone on all four ERP components, although the effect reached a greater level of significance for later components. The main effect of tone on the N1 measured from 100-150 msec was $F(2, 58) = 2.81$, $p < 0.1$, the main effect of tone on the P2 measured from 150-230 msec was $F(2, 58) = 5.50$, $p < 0.01$, the main effect of tone on the N2 was $F(2, 58) = 7.44$, $p < 0.001$, and the main effect of tone on the P2 was $F(2, 58) = 3.22$, $p < 0.05$.

The effects of target frequency and task performance on ERP latency were again more limited than the effects on amplitude. While there was no target and tone interaction in the N1 component, there was a significant interaction between target, tone, and predictability ($F(4, 118) = 2.42$, $p < 0.05$). Interactions between tone frequency and target frequency were only significant for P2 ($F(4, 118) = 2.13$, $p < 0.1$), which also had a main effect of tone ($F(2, 58) = 8.36$, $p < 0.001$). As in Experiment 2, there was a main effect of accuracy on the latency of the N1 component ($F(1, 28) = 9.21$, $p < 0.01$), in addition to an interaction between accuracy and predictability ($F(1, 28) = 3.25$, $p < 0.1$). There was also a main effect of predictability on the N1 ($F(1, 28) = 24.81$, $p < 0.001$). Predictability also had a main effect on the P2 ($F(1, 28) = 17.17$, $p < 0.001$) and an interaction effect with accuracy on the P3b ($F(1, 28) = 4.46$, $p < 0.05$).

5.4 Discussion

The present study replicated the main effects of task performance on FFR amplitude. High-performing participants exhibited reduced FFR amplitudes relative to low performers. Additionally, no differences in phase-locking value (PLV) were observed between performance groups. These results are consistent with previous findings linking efficient auditory processing to increased neural selectivity and suggesting that FFR amplitude is not a direct index of

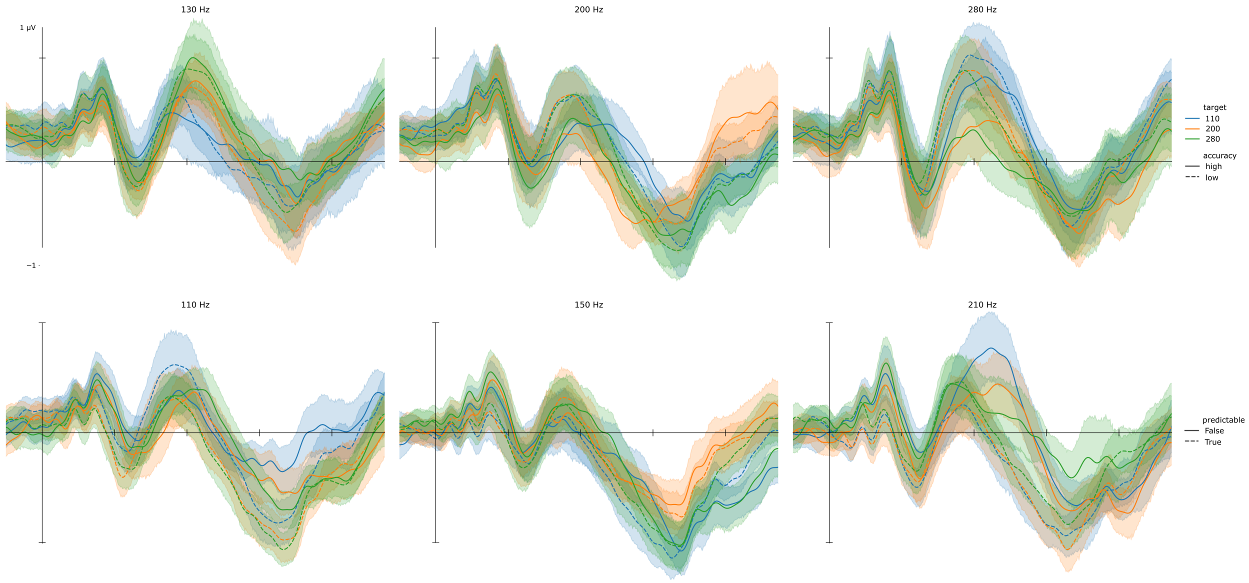


Figure 5.5: ERP amplitudes to the stimulus tones were significantly different if the tones were target tones. ERP amplitudes also varied depending on the predictability of the tone (middle) and on task performance (bottom). A small but significant four-way interaction was found between tone frequency, target frequency, task performance, and predictability.

the fidelity of auditory encoding. Contrary to predictions from some attention and predictive coding models, stimulus predictability did not modulate FFR amplitude or consistency, indicating that regularity in tone sequences may not influence subcortical encoding under these task conditions. In contrast, strong effects of both task performance and task demands were found in ERPs, particularly in the later components such as the P3b, suggesting that while task demands influence late cortical responses, their effects on earlier auditory encoding may be limited.

The absence of FFR amplitude differences between predictable and unpredictable active conditions suggests that the general, non-selective neural recruitment often observed during passive listening does not extend to very easy active tasks. This finding supports the idea that passive and active listening may engage distinct neurobiological mechanisms, even when the attentional demands of the active task are minimal. The current model proposes that passive listening elicits broad increases in FFR amplitude due to a lack of attentional suppression. Although the task used in this experiment was easy, with most participants at the performance ceiling, the observed FFR response was not different from the FFR response to the more difficult active task. This suggests that easy active tasks still recruit dynamic tuning mechanisms that enhance selectivity by suppressing irrelevant neural responses. However, the current model also proposes that effort increases FFR amplitudes, as suggested in the previous experiments. The lack of increased FFR amplitudes in the unpredictable active condition relative to the predictable condition does not support this prediction. Unfortunately, the lack of a measurable FFR amplitude difference across conditions makes it difficult to determine whether the different neural mechanisms were engaged under different task demands, or whether they exerted opposing or canceling effects that were not captured by our dependent measures. As with any null result, it remains unclear whether the observed similarity reflects engagement of the same underlying mechanism or a convergence of different mechanisms yielding similar output.

Despite these inconclusive results, this experiment remains valuable for highlighting the challenges of disentangling the overlapping effects of expertise, attention, and effort during audition. It raises the possibility that the current task manipulations and measures may lack the sensitivity to detect subtle shifts in neural encoding strategies, as previous task demands, such as target identity and a visual distractor task, did not elicit any changes in FFR amplitudes. Additionally, individual differences in task responses may have obscured differences in underlying neural activity. The previous experiments presented in this dissertation have shown that individual differences in effort and expertise may modulate the mechanisms of neural encoding and auditory attention recruited. Splitting participants up into groups based on task performance trajectory enabled the comparison between the independent effects of effort and expertise. However, because block types were alternated in the current experiment, it was not possible to analyze performance trajectories in the same way, making it more difficult to dissociate the effects of effort from efficiency. These findings underscore the need for continued replication and methodological improvement to build a model of auditory attention that is robust and explains findings across different behavioral and neuroimaging data.

CHAPTER 6

GENERAL DISCUSSION

The overarching goal of this dissertation was to contribute to the development of a more comprehensive neurobiological account of auditory attention by building a cognitive model of the role of early, phase-locked neural responses in audition and auditory attention. As past researchers such as Fritz and colleagues (2007) and Holt and colleagues (2018) argue, current models of auditory attention are often shaped by the methodological constraints of the data that inform them. Models derived from human neuroimaging tend to emphasize slow, broad modulations in activity associated with attentional control, and support the assumption that enhancements in sensory encoding are implemented by increases in neural gain. In contrast, models grounded in animal electrophysiology, which focus on fast, precise changes in neural tuning at the level of single units, propose that auditory attention is instantiated by changes in the selectivity of neural activity. Due to the mixed literature on the effects of top-down attention and the limited research on the effects of experience on the FFR under active listening conditions, it remains unknown whether attention and long-term experience modulated the FFR through changes in gain or selectivity. The literature has instead defaulted to the assumption that greater FFR amplitudes necessarily reflect increased encoding fidelity. The FFR is a scalp-recorded signal that captures phase-locked neural activity across cortical and subcortical levels of the auditory system (Holmes et al., 2018; Lerud et al., 2023; Coffey et al., 2016, 2017; Tichko & Skoe, 2017) and enables the high-temporal-resolution measurement of early auditory encoding in humans. Studying the changes in gain and selectivity in the FFR under different attention demands in participants with a range of domain-specific expertise in this dissertation has helped develop a model of how phase-locked neural firing indexed by the FFR helps carry out audition and auditory attention. The data from the studies presented suggest that domain-specific expertise can shift the baseline selectivity of phase-locked responses, and flexibly recruit cognitive resources to synchronize to periodic stimulus features under effortful listening conditions.

Overview of Findings

Experiment 1 employed an intermodal attention task designed as a conservative replication of prior FFR studies reporting attentional modulation. Unlike previous studies, this paradigm controlled for important confounds such as stimulus length and type, and incorporated a continuous measure of task performance to assess whether attentional effects on FFR amplitude varied with behavioral outcomes. While prior literature has demonstrated increased FFR amplitude under conditions of selective attention, no main effect of attentional condition was observed in this experiment. Instead, a significant increase in FFR amplitude was found in high-performing participants, with no corresponding change in phase-locking value (PLV). These results suggest that FFR amplitude may be more sensitive to individual differences in task engagement or effort, rather than the attentional condition. Importantly, the absence of PLV differences implies that changes in FFR amplitude are not driven by enhanced phase consistency but may reflect increased neural recruitment or gain. This study laid the groundwork for the dissociation between effort and selectivity that was further explored in subsequent experiments.

Experiment 2 expanded on these findings by developing a new task paradigm designed to systematically examine the interaction between task performance and task demands. The oddball-style auditory task used here was simpler than that of Experiment 1, with a more lenient response criterion that provided positive feedback even for approximate answers. In this easier task context, the direction of the task performance effect reversed, with high performers exhibiting decreased FFR amplitudes compared to low performers. Again, PLVs remained unchanged, reinforcing the notion that FFR amplitude and phase consistency index different aspects of neural encoding. These findings support the hypothesis that under low-effort conditions, high performers may rely on more efficient, selective neural tuning mechanisms that reduce overall response magnitude. Additionally, while FFR measures remained stable across task demands, late auditory ERPs (e.g., N1, N2, P3b) showed sensitivity to task variables such as target identity. This dissociation highlights a limit in the top-down effects of task demands, suggesting that the impacts of immediate context on neural responses are limited to late, large-amplitude cortical responses.

The next experiment is the most important experiment presented in this dissertation. Experiment 3 served as the pivotal test of competing models of auditory attention, specifically it compared whether gain modulation or dynamic tuning better accounts for neural responses under active and passive listening, and grouped participants based on performance trajectory to test predictions about the effects of effort. Gain-based models predict enhanced FFR amplitudes during effortful active listening, whereas dynamic tuning models anticipate greater selectivity and reduced response magnitude in skilled listeners. This experiment successfully replicated the main findings from Experiment 2, finding again that high task performers exhibited decreased FFR amplitudes with no change in PLV. Additionally, this study found that passive listening significantly increased FFR amplitudes across all participant groups. This result provides strong evidence against the assumption that higher FFR amplitudes necessarily indicate superior encoding, and instead suggests that passive listening may trigger broadly distributed, non-selective auditory responses. To further disentangle the role of effort, participants were grouped based on their performance trajectories. Consistent high performers showed the lowest FFR amplitudes, while "improvers" those who increased in accuracy over time, exhibited higher FFR amplitudes than "decliners" despite similar average performance. These results suggest that improvers were engaging more effortful listening strategies, leading to increased neural gain. The findings from Experiment 3 not only reinforce the dissociation between effort and selectivity but also provide a compelling framework for understanding how expertise and task engagement differentially modulate early auditory encoding.

The last experiment of this dissertation sought to test the limits of this emerging framework by using a novel variant of the active listening task that imposed minimal attentional demands. In this version, tone sequences followed a predictable repeating pattern (e.g., "AABBCC..."), rendering the task trivially easy and pushing attentional demands toward the lower bound of active listening. The purpose was to investigate whether very low-effort active tasks would elicit neural responses resembling those seen in passive listening, characterized by increased FFR amplitudes. However, the results did not support this hypothesis. FFR amplitudes in the easy predictable condition did not increase to levels observed during passive listening, nor did they significantly differ from FFR responses during the standard active task. While these null findings make interpretation difficult, they suggest that passive listening and very low-effort active listening are not neurally equivalent. It remains unclear whether the same attentional mechanisms were engaged in both tasks or whether different mechanisms produced overlapping effects. Despite these ambiguities, Experiment 4 underscores the complexity of dissociating the roles of attention, effort, and predictability in shaping early auditory responses. It also highlights the need for continued replication and finer-grained manipulations to clarify the conditions under which different neurobiological mechanisms of auditory attention are engaged.

The role of expertise, attention, and effort in engaging different mechanisms of auditory encoding

Across Experiments 2, 3, and 4, high-performing individuals consistently exhibited reduced FFR amplitudes during the easy selective auditory attention tasks. This result supports the idea that efficient processing does not require a stronger neural response but may instead reflect a more selective, streamlined use of neural resources. Experiment 3 extended this observation by showing similarly reduced FFR amplitudes in high performers even under passive listening conditions. This finding suggests that the increased selectivity in auditory encoding associated with expertise may be global and context-independent, persisting even when attention is not explicitly directed toward the stimulus. Moreover, across experiments, FFR amplitudes changes associated with task performance did not interact with other factors such as predictability, target identity, or task modality. This independence supports a model of expertise-driven selectivity that operates broadly across listening conditions, one in which experience leads to stable, efficient neural tuning mechanisms.

Across all studies, task performance was strongly, positively correlated with years of musical experience, and the accuracy of high performers were near ceiling, providing support for the interpretation of performance as a strong correlate for auditory expertise. Additionally, in Experiment 3, consistently high performers had significantly more musical experience than participants in the improver or decliner groups, despite showing no differences in auditory working memory (as measured by the auditory N-back task). This dissociation suggests that musical expertise does not simply reflect generalized cognitive abilities but may instead reflect domain-specific auditory tuning. Long-term musical training has been shown to induce structural and functional changes in auditory and motor regions (Gaser & Schlaug, 2003; Münte et al., 2002; Schlaug, 2001), including enhanced myelination, cortical thickness, and white matter connectivity in relevant networks. Under non-challenging conditions, these neural adaptations may allow experts to process auditory information more selectively, resulting in reduced FFR amplitudes without compromising temporal precision. Indeed, across all experiments, reductions in FFR amplitude among experts were not accompanied by reductions in PLV, suggesting that temporal consistency of the response was preserved even as overall response magnitude decreased. This pattern aligns with predictions from dynamic tuning models of attention, in which experience enables the auditory system to sharpen its responses to relevant stimuli by suppressing irrelevant neural activity.

Neurophysiological support for such selective tuning mechanisms comes from animal models of auditory and visual attention. In the auditory domain, Schwartz and David (2018) recorded intracranial activity from mice exposed to two competing tone streams and found that attention to a specific frequency band led to a suppression of spike rates to irrelevant stimuli, without enhancing the response to attended probe tones. This suggests that attention operates not by boosting all relevant signals but by suppressing distractors, a hallmark of efficient encoding. Similarly, in the visual system, Chelazzi and colleagues (1993) reported that neurons in monkey visual cortex responded less to non-target stimuli even when those stimuli contained features to which the neurons were tuned. Subsequent studies have shown that with familiarity and training, neural responses become increasingly sparse and selective (e.g., Reynolds & Chelazzi, 2004). This growing selectivity, also described within the framework of biased competition (Desimone & Duncan, 1995), may reflect the same principles at work in auditory expertise, whereby training leads to a refinement of receptive fields and a reduction in unnecessary neural activity. The reduced FFR amplitudes observed in expert listeners in this dissertation, particularly in low-effort contexts, likely reflect an enhancement in the precision of sensory processing by tuning out irrelevant noise rather than increasing overall response gain.

The findings across the studies in this dissertation also point toward a distinct role for cognitive effort in modulating neural gain during auditory encoding. This interpretation is supported most directly by the increase in FFR amplitudes observed among high performers in the challenging intermodal task of Experiment 1, a result which contrasts with the decreased FFR amplitudes seen in high performers in the easier auditory task of the subsequent experiments. The divergent results across tasks with different cognitive demands suggest that the magnitude of the FFR may reflect not only sensory encoding quality but also the degree of cognitive effort being exerted. Additional evidence comes from Experiment 3, where participants were grouped by their performance trajectories. “Improvers”, who began the task with lower performance but showed consistent gains, exhibited significantly larger FFR amplitudes than both consistent high performers and decliners. Given that improvers had slightly lower auditory working memory abilities than the other groups but less musical experience than high performers, their enhanced FFR amplitudes may reflect increased effort rather than pre-existing cognitive or training-based advantages. Together, these findings suggest that expending greater cognitive effort in the service of task performance enhances the magnitude of auditory encoding, consistent with the idea of effort-induced gain modulation. These results align well with the Framework for Understanding Effortful Listening (FUEL) proposed by Pichora-Fuller and colleagues (2016), which posits that listening effort arises from the dynamic allocation of cognitive resources based on task demands, motivation, and listener capacity. According to FUEL, individuals will engage greater effort and more cognitive resources when tasks are challenging but achievable, particularly when motivation to perform well is high. The FFR amplitude increases observed in this dissertation appear to reflect such effort-driven resource recruitment.

In addition to associations between FFR amplitudes and task performance, Experiment 3 found an overall increase in FFR amplitudes across all performance groups during passive listening. This main effect was accompanied by weak interaction effects between performance group or performance trajectory, and task condition, suggesting that increases in FFR amplitudes under passive listening may be lower in high task performers. These results suggest that, in comparison to active listening tasks, passive listening does not require the suppression of any part of the stimulus stream. Under passive listening, neural activity can be freely patterned by oncoming stimuli and bottom-up auditory activity, producing greater neural responses with less selectivity. Additionally, the lack of any significant difference in PLV between conditions and task performance groups suggests that the consistency of phase locking is stable across conditions, effort, and expertise and that it is the degree of neural recruitment involved that changes. The interactions between performance group and passive listening, in conjunction, indicate that the auditory systems of high performers appear to be shaped by long-term experience in a way that promotes efficient and selective processing, which generalizes beyond specific task contexts.

An alternative explanation is that the specific active task used, and not active tasks generally, requires a more selective response. The task used in this experiment required participants to monitor only a specific frequency band for targets and match incoming stimuli to a mental template of target tones. This response specificity potentially activates a more selective response. The current paradigm does not allow for the two accounts to be directly compared, but both interpretations support the conclusion that passive listening may recruit a broader, more non-selective response. The two interpretations posit different cognitive roles for the decrease in FFR amplitude. A way to disambiguate whether it is active listening that is causing the decrease in FFR amplitude or the specific task that is engaging a more selective response is to compare, within-subjects, FFRs on the current active task and an active task that requires the monitoring of a broader range of frequencies to potentially engage a larger neural population.

Lastly, we found no effects of target tone identity, intermodal attention, or predictability on FFR amplitudes. These findings suggest that immediate task context and selective attention manipulations may exert minimal influence on phase-locked brainstem activity, a conclusion that echoes the historically inconsistent effects of attention on FFRs reported in the literature (e.g., Galbraith & Arroyo, 1993; Varghese et al., 2015). However, the lack of task demand effects, with the exception of passive listening, is somewhat concerning. Prominent models of auditory attention (e.g., Shinn-Cunningham, 2008; Fritz et al., 2007) generally do not distinguish passive listening as a categorically different mechanism from other selective attention tasks, raising questions about the nature of the processes engaged under passive conditions. The consistent pattern of null effects for more fine-grained task manipulations cautions against introducing ad hoc mechanisms to explain isolated findings, underscoring the need for rigorous replication and theoretical restraint. If the current results are reliable, they imply that passive listening may indeed engage a fundamentally different mode of auditory processing, one that is more diffuse and less suppressive than the mechanisms recruited under active task engagement. These results may also suggest that all forms of active listening may engage auditory attention, but that differences between types of attentional tasks may be too subtle or occur at later processing stages to be captured by changes in the FFR. This interpretation is supported by evidence of modulations in the N1, N2 and P3b ERP components by task demands, suggesting that selective attention and contextual effects may emerge more robustly in late cortical auditory processing.

The findings presented in this dissertation provide evidence that FFRs appear to index changes in neural gain and selectivity associated with task engagement, experience, and effort, rather than simply reflecting encoding fidelity. Previous models of auditory attention, informed by experiments reporting enhanced BOLD signals in frequency-selective auditory regions under attentive conditions (e.g., Alho et al., 2003; Hill & Miller, 2010; Paltoglou et al., 2009; Luthra et al., 2024) have assumed that auditory attention is supported solely by increased neural gain. In comparison, dynamic tuning and biased competition models based on intracranial recordings in animals have proposed that neural activity becomes more selective with attention and experience (Desimone & Duncan, 1995; Fritz et al., 2007). Here we find that high performers had decreased FFR amplitudes relative to low performers during easy tasks and passive listening, and increased FFR amplitudes relative to low performers during a more difficult task. These results indicate that dynamic tuning shifts the baseline response of performers with musical experience or expertise, rendering their responses more selective with greater suppression of irrelevant activity. Relatedly, these results suggest that experts may more flexibly increase neural gain during difficult tasks. In sum, the results here suggest that gain modulation and selectivity may both operate under different contexts, modulated by effort and expertise respectively, a view that integrates and refines existing models of auditory attention and processing.

Implications for mechanistic models of auditory encoding

The findings presented in this dissertation offer new insights into the neural mechanisms of auditory attention by examining changes in the FFR. A key question concerning the neuroanatomical locus of the observed amplitude change should be addressed. The FFR has traditionally been assumed to primarily be a subcortical response, if this is the case, then the results in this dissertation might not correspond directly to attentional modulations observed in cortical fMRI or single- and multi-unit studies (e.g., Fritz et al., 2003; Shinn-Cunningham, 2008). However, there is strong evidence to suggest that the FFR is not exclusively subcortical. While traditionally attributed to brainstem nuclei, such as the inferior colliculus and cochlear nucleus, recent work has shown cortical contributions, especially at lower fundamental frequencies (F0s) and in response to complex sounds (e.g., Coffey et al., 2016, 2017, 2019; Tichko & Skoe, 2017; Lerud et al., 2023). This suggests that both subcortical and cortical sources contribute to the scalp-measured FFR, and that their relative contributions may vary by frequency, task, and stimulus. The observed results are therefore likely to have a bearing on these cortical models of auditory encoding. Furthermore, observing attentional or effort-related changes in the FFR suggests that some top-down influence is modulating early encoding. If this modulation happens at both cortical and subcortical structures, it challenges cortical-centric models of auditory attention and argues for more hierarchical or distributed frameworks (e.g., Holt & Lotto, 2008; Suga et al., 2002). Overall, even if the locus of FFR effects is not fully resolved, the findings presented here suggest plasticity and flexibility in highly precise, phase-locked responses across the auditory pathway.

More critical than the discussion of the locus of attention effects on the FFR is the tension between models of neural encoding that are symbolic, or tonotopic and feature-based, and models that are temporal and rely on precise phase locking. The frequency-following response (FFR) reflects phase-locked neural firing to periodicities in sound, particularly the temporal envelope and fine structure of stimuli. As such, it is fundamentally a time-domain representation of sound, preserving both stimulus frequency content and temporal regularity (Skoe & Kraus, 2010). However, many dominant models of auditory attention, such as gain modulation and dynamic tuning frameworks, are based on symbolic or feature-based representations, where neurons are treated as selective channels for specific acoustic attributes (e.g., frequency, spatial location, pitch) and where attention modulates response magnitude or sharpens receptive field properties.

The full reconciliation between these representational assumptions is beyond the scope of this dissertation, but a discussion is necessary for a realistic estimate of the bearing of this data on these traditional, tonotopic or feature-based, models of auditory attention. It has been suggested that symbolic and temporal encoding schemes are not mutually exclusive, but may instead be complementary dimensions of auditory representation (Joris & Schreiner, 2004; Oxenham, 2013; Downer et al., 2021; Wang et al., 2008). Temporal coding through phase-locked firing may dominate early in the auditory pathway, particularly as studies have shown that phase locking is limited to lower frequencies (<200 Hz) in cortical areas (Holmes et al., 2018; Bidelman et al., 2018). In comparison, phase locked activity in neurons in the auditory cortex may encode more slowly varying acoustic signals (such as the envelope of speech or rhythms) (Schroeder and Lakatos, 2009; Luo and Poeppel, 2007; Doelling and Poeppel, 2015; Ding and Simon, 2014), while rapid or static features are encoded via changes in average firing rates. This suggests that temporal precision at early subcortical stages gradually transitions into a greater reliance on rate-based or spectral feature representations in cortex, enabling efficient detection and categorization of sounds. In the context of the current data, the amplitude of the FFR can be thought of as reflecting not just how strongly neurons are firing, but how many neurons are synchronously firing in phase with the stimulus. A decrease in FFR amplitude in high performers may reflect increased selectivity, but with greater tuning to the relevant frequency. This is compatible with dynamic tuning models, which predict more focused neural engagement with expertise (Fritz et al., 2003;

Schwartz & David, 2018), and with gain modulation models, where effortful tasks broadly increase response magnitude (Pichora-Fuller et al., 2016). Both models assume a symbolic architecture, but their effects can manifest in changes to temporal encoding measures.

Methodological Considerations and Null Effects

The experiments presented in this dissertation have raised a few methodological considerations that should be considered for future FFR and EEG-FFR experiments. Experiment 1 demonstrated that short stimulus signals, containing $<$10 cycles of the frequency of interest, result in low temporal resolutions in the resulting power estimates of the evoked FFR signals. The use of short stimulus durations, many with only ten or fewer cycles of the stimulus pitch, may explain the mixed results observed in FFR studies on attention effects.

Additionally, all studies in this dissertation used single-polarity stimuli to preserve the full phase-locked response to the temporal fine structure of the auditory stimulus. While alternating polarities are often employed to suppress the cochlear microphonic and isolate envelope-following responses, single-polarity presentations allow for more precise measurement of the neural response to fine-structure components, particularly when analyzing responses in the frequency domain. This approach is supported by evidence from Bidelman and colleagues (2015), who demonstrated that common-average referencing (CAR), commonly used in EEG experiments, selectively reduces the contribution of subcortical sources to the FFR, effectively biasing the recorded signal toward cortical activity. The use of single-polarity stimuli was used to preserve both cortical and subcortical contributions to the signal (Bidelman et al., 2015; Coffey et al., 2019). Moreover, single-polarity stimuli are increasingly common in FFR studies that prioritize frequency-domain analyses (e.g., Tichko & Skoe, 2017), particularly those aimed at quantifying frequency-specific response amplitudes and phase locking consistency. This methodological choice aligns with emerging recommendations in the literature for characterizing the full spectral profile of FFRs without artificially attenuating key components of the response.

Another consideration is the use of task performance as a proxy for expertise. Although past studies on the effects of experience and learning on neural representations of stimulus features have used task performance as a measure of expertise, there is a reasonable argument to be made that task performance does not solely reflect expertise. The FUEL framework argues that performance and neural engagement may be independently shaped by effort and expertise (Pichora-Fuller et al., 2016). Rönnberg and colleagues (2008) demonstrate that listeners may show identical performance across tasks yet expend vastly different cognitive resources as measured by reaction times, self-report, and physiological markers. To complicate the relationship between task performance and expertise, multiple studies on reward conditions and task complexity have shown that these factors may influence the extent to which performance reflects effort and expertise in each unique task (Gagné et al., 2017; Dewaele et al., 2025). These complications are the reason why participants were split into three separate groups based on task performance trajectories in Experiment 3, and only the electrophysiological responses of participants who perform consistently at ceiling are considered indicative of expertise. Future experiments on the effects of expertise and effort on auditory encoding should consider adding measures of pupillary dilation, skin conductance, and self-report measures of listening effort (Bishop et al., 2021; Alhanbali et al., 2019; Winn et al., 2015). Indeed, Endestad and colleagues (2020) compared expert pianists with non-musicians using pupillometry and fMRI and found that pianists exhibited lower pupil dilation and reduced neural effort when listening to music, suggesting experts use fewer resources to achieve high performance.

Finally, whereas most FFR studies use either amplitude or PLV as measures of the general strength of phase locking, experiments in this dissertation compared the two measures as indexing complementary but separate components of the FFR. Increases in amplitude do not necessarily mean increases in PLV. Power or amplitude reflects the overall magnitude of the neural response at a given frequency, and as with ERPs, this is generally considered to reflect the number of neurons responding synchronously in the summed scalp potential. In contrast, PLV quantifies the consistency of the phase of the response across trials, in other words how reliably neural activity entrains in time to the stimulus. While both measures can be influenced by stimulus characteristics or cognitive factors, such as attention or expertise, they do not necessarily covary. A neural response can be strong but inconsistent in timing or weak but precisely timed. Thus, jointly examining PLV and power allows for a more complete characterization of the magnitude and reliability of auditory encoding.

Conclusion

Potential follow-up experiments for the work in this dissertation include replications of current experiment results with passive listening paradigms used in other experiments, such as recording FFRs to the synthesized vowel /da/ commonly used in other experiments (e.g., Tierney et al., 2013; 2015; Russo et al., 2005; Krizman et al., 2012; Skoe et al., 2017). The present data also suggest that the neural mechanisms involved in encoding stimuli, including gain modulation and dynamic tuning, depend heavily on individual differences in musical experience and effort. Future experiments should directly survey participants on the amount of cognitive effort they expend on studies, perhaps through behavioral measures or measures such as pupillary dilation or skin conductance. A within-subjects study comparing FFRs during tasks of varying difficulty would also help assess the proposed effects of effort in engaging gain modulation mechanisms. Finally, including a longitudinal study on the effects of training on task performance and FFRs could provide valuable insight into the timescale of plasticity in early auditory encoding.

Taken together, the findings presented in this dissertation offer new insights into how auditory encoding is shaped by attention, effort, and expertise. By using the frequency-following response to examine early, phase-locked neural activity, these studies demonstrate that different neurobiological mechanisms, including gain modulation and dynamic tuning, are engaged depending on task demands and individual listener characteristics. Importantly, the work highlights that higher FFR amplitude does not uniformly indicate enhanced encoding, but may instead reflect increased neural gain in response to cognitive effort. At the same time, decreased FFR amplitudes in experts suggest a shift toward more selective and efficient processing. These results underscore the importance of considering both contextual and individual-level factors when interpreting neural measures of auditory attention and point toward the need for more integrative models that can reconcile symbolic and temporal accounts of auditory representation. Ultimately, this work advances our understanding of how the brain flexibly adapts its encoding strategies in response to the varying demands of the auditory environment.

REFERENCES

Alhanbali, S., Dawes, P., Millman, R. E., & Munro, K. J. (2019). Measures of listening effort are multidimensional. *Ear and Hearing*, *40*(5), 1084–1097. https://doi.org/10.1097/aud.0000000000000697

Alho, K. (1992). Selective attention in auditory processing as reflected by event-related brain potentials. *Psychophysiology*, *29*(3), 247–263. https://doi.org/10.1111/j.1469-8986.1992.tb01695.x

Alho, K., Salonen, J., Rinne, T., Medvedev, S. V., Hugdahl, K., & Hämäläinen, H. (2012). Attention-related modulation of auditory-cortex responses to speech sounds during dichotic listening. *Brain Research*, *1442*, 47–54. https://doi.org/10.1016/j.brainres.2012.01.007

Alho, K., Töttölä, K., Reinikainen, K., Sams, M., & Näätänen, R. (1987). Brain mechanism of selective listening reflected by event-related potentials. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *68*(6), 458–470. https://doi.org/10.1016/0168-5597(87)90057-8

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bharadwaj, H. M., Lee, A. K. C., & Shinn-Cunningham, B. G. (2014). Measuring auditory selective attention using frequency tagging. *Frontiers in Integrative Neuroscience*, *8*. https://doi.org/10.3389/fnint.2014.00006

Bidelman, G. M. (2015). Multichannel recordings of the human brainstem frequency-following response: Scalp topography, source generators, and distinctions from the transient abr. *Hearing Research*, *323*, 68–80. https://doi.org/10.1016/j.heares.2015.01.011

Bidelman, G. M. (2018). Subcortical sources dominate the neuroelectric auditory frequency-following response to speech. *NeuroImage*, *175*, 56–69. https://doi.org/10.1016/j.neuroimage.2018.03.060

Bidelman, G. M., & Alain, C. (2015). Musical training orchestrates coordinated neuroplasticity in auditory brainstem and cortex to counteract age-related declines in categorical vowel perception. *The Journal of Neuroscience*, *35*(3), 1240–1249. https://doi.org/10.1523/jneurosci.3292-14.2015

Bidelman, G. M., & Momtaz, S. (2021). Subcortical rather than cortical sources of the frequency-following response (ffr) relate to speech-in-noise perception in normal-hearing listeners. *Neuroscience Letters*, *746*, 135664. https://doi.org/10.1016/j.neulet.2021.135664

Bigdely-Shamlo, N., Mullen, T., Kothe, C., Su, K.-M., & Robbins, K. A. (2015). The prep pipeline: Standardized preprocessing for large-scale eeg analysis. *Frontiers in Neuroinformatics*, *9*. https://doi.org/10.3389/fninf.2015.00016

Binder, J. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cerebral Cortex*, *10*(5), 512–528. https://doi.org/10.1093/cercor/10.5.512

Bishop, L., Jensenius, A. R., & Laeng, B. (2021). Musical and bodily predictors of mental effort in string quartet music: An ecological pupillometry study of performers and listeners. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.653021

Blank, H., & Davis, M. H. (2016). Prediction errors but not sharpened signals simulate multivoxel fmri patterns during speech perception (R. Zatorre, Ed.). *PLOS Biology*, *14*(11), e1002577. https://doi.org/10.1371/journal.pbio.1002577

Chelazzi, L., Miller, E. K., Duncan, J., & Desimone, R. (1993). A neural basis for visual search in inferior temporal cortex. *Nature*, *363*(6427), 345–347. https://doi.org/10.1038/363345a0

Christopher, S. (2011). Expectation and attention in visual cognition: Neural and computational approaches. *Frontiers in Human Neuroscience*, *5*. https://doi.org/10.3389/conf.fnhum.2011.207.00036

Churchland, S., P. S. (2016). In *The computational brain*. The MIT Press. https://doi.org/10.7551/mitpress/11207.003.0006

Coch, D., Sanders, L. D., & Neville, H. J. (2005). An event-related potential study of selective auditory attention in children and adults. *Journal of Cognitive Neuroscience*, *17*(4), 605–622. https://doi.org/10.1162/0898929053467631

Coffey, E. B. J., Herholz, S. C., Chepesiuk, A. M. P., Baillet, S., & Zatorre, R. J. (2016). Cortical contributions to the auditory frequency-following response revealed by meg. *Nature Communications*, *7*(1). https://doi.org/10.1038/ncomms11070

Coffey, E. B. J., Nicol, T., White-Schwoch, T., Chandrasekaran, B., Krizman, J., Skoe, E., Zatorre, R. J., & Kraus, N. (2019). Evolving perspectives on the sources of the frequency-following response. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-019-13003-w

Coffey, E. B., Musacchia, G., & Zatorre, R. J. (2017). Cortical correlates of the auditory frequency-following and onset responses: Eeg and fmri evidence. *Journal of Neuroscience*, *37*(4), 830–838.

Crowley, K. E., & Colrain, I. M. (2004). A review of the evidence for p2 being an independent component process: Age, sleep and modality. *Clinical Neurophysiology*, *115*(4), 732–744. https://doi.org/10.1016/j.clinph.2003.11.021

David, S. V., Fritz, J. B., & Shamma, S. A. (2012). Task reward structure shapes rapid receptive field plasticity in auditory cortex. *Proceedings of the National Academy of Sciences*, *109*(6), 2144–2149. https://doi.org/10.1073/pnas.1117717109

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*(1), 193–222. https://doi.org/10.1146/annurev.ne.18.030195.001205

Dewaele, L., Cardinaels, E., & Van den Abbeele, A. G. H. L. (2025). Reward choices: Experimental evidence on cognitive task performance. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.5244883

Dick, F. K., Lehet, M. I., Callaghan, M. F., Keller, T. A., Sereno, M. I., & Holt, L. L. (2017). Extensive tonotopic mapping across auditory cortex is recapitulated by spectrally-directed attention, and systematically related to cortical myeloarchitecture. *BioRxiv*. https://doi.org/10.1101/142513

Dien, J., Michelson, C. A., & Franklin, M. S. (2010). Separating the visual sentence n400 effect from the p400 sequential expectancy effect: Cognitive and neuroanatomical implications. *Brain Research*, *1355*, 126–140. https://doi.org/10.1016/j.brainres.2010.07.099

Ding, N., & Simon, J. Z. (2014). Cortical entrainment to continuous speech: Functional roles and interpretations. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00311

Doelling, K. B., & Poeppel, D. (2015). Cortical entrainment to music and its modulation by expertise. *Proceedings of the National Academy of Sciences*, *112*(45). https://doi.org/10.1073/pnas.1508431112

Downer, J. D., Verhein, J. R., Rapone, B. C., O'Connor, K. N., & Sutter, M. L. (2020). An emergent population code in primary auditory cortex supports selective attention to spectral and temporal sound features. *BioRxiv*. https://doi.org/10.1101/2020.03.09.984773

Endestad, T., Godøy, R. I., Sneve, M. H., Hagen, T., Bochynska, A., & Laeng, B. (2020). Mental effort when playing, listening, and imagining music in one pianist's eyes and brain. *Frontiers in Human Neuroscience*, *14*. https://doi.org/10.3389/fnhum.2020.576888

Engelmann, J. B. (2009). Combined effects of attention and motivation on visual task performance: Transient and sustained motivational effects. *Frontiers in Human Neuroscience*, *3*. https://doi.org/10.3389/neuro.09.004.2009

Folstein, J. R., & Van Petten, C. (2007). Influence of cognitive control and mismatch on the n2 component of the erp: A review. *Psychophysiology*, *45*(1), 152–170. https://doi.org/10.1111/j.1469-8986.2007.00602.x

Friston, K., Price, C., Fletcher, P., Moore, C., Frackowiak, R., & Dolan, R. (1996). The trouble with cognitive subtraction. *NeuroImage*, *4*(2), 97–104. https://doi.org/10.1006/nimg.1996.0033

Fritz, J. B., Elhilali, M., David, S. V., & Shamma, S. A. (2007). Auditory attention—focusing the searchlight on sound. *Current Opinion in Neurobiology*, *17*(4), 437–455. https://doi.org/10.1016/j.conb.2007.07.011

Gagné, J.-P., Besser, J., & Lemke, U. (2017). Behavioral assessment of listening effort using a dual-task paradigm: A review. *Trends in Hearing*, *21*. https://doi.org/10.1177/2331216516687287

Galbraith, G. C., & Arroyo, C. (1993). Selective attention and brainstem frequency-following responses. *Biological Psychology*, *37*(1), 3–22. https://doi.org/10.1016/0301-0511(93)90024-3

Galbraith, G. C., Bhuta, S. M., Choate, A. K., Kitahara, J. M., & Mullen, T. A. (1998). Brain stem frequency-following response to dichotic vowels during attention. *NeuroReport*, *9*(8), 1889–1893. https://doi.org/10.1097/00001756-199806010-00041

Galbraith, G. C., & Brown, W. S. (1990). Cross-correlation and latency compensation analysis of click-evoked and frequency-following brain-stem responses in man. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *77*(4), 295–308. https://doi.org/10.1016/0168-5597(90)90068-o

Galbraith, G. C., & Doan, B. Q. (1995). Brainstem frequency-following and behavioral responses during selective attention to pure tone and missing fundamental stimuli. *International Journal of Psychophysiology*, *19*(3), 203–214. https://doi.org/10.1016/0167-8760(95)00008-g

Galbraith, G. C., & Kane, J. M. (1993). Brainstem frequency-following responses and cortical event-related potentials during attention. *Perceptual and Motor Skills*, *76*, 1231–1241. https://doi.org/10.2466/pms.1993.76.3c.1231

Galbraith, G. C., Olfman, D. M., & Huffman, T. M. (2003). Selective attention affects human brain stem frequency-following response. *NeuroReport*, *14*(5), 735–738. https://doi.org/10.1097/00001756-200304150-00015

Gaser, C., & Schlaug, G. (2003). Brain structures differ between musicians and non-musicians. *The Journal of Neuroscience*, *23*(27), 9240–9245. https://doi.org/10.1523/jneurosci.23-27-09240.2003

Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2002, September). Expertise for cars and birds recruits brain areas involved in face recognition. In *Foundations in social neuroscience* (pp. 277–292). The MIT Press. https://doi.org/10.7551/mitpress/3077.003.0022

Gauthier, I., Williams, P., Tarr, M. J., & Tanaka, J. (1998). Training 'greeble' experts: A framework for studying expert object recognition processes. *Vision Research*, *38*(15–16), 2401–2428. https://doi.org/10.1016/s0042-6989(97)00442-2

Goh, H.-T., Gordon, J., Sullivan, K. J., & Winstein, C. J. (2014). Evaluation of attentional demands during motor learning: Validity of a dual-task probe paradigm. *Journal of Motor Behavior*, *46*(2), 95–105. https://doi.org/10.1080/00222895.2013.868337

Golestani, N., & Zatorre, R. J. (2004). Learning new sounds of speech: Reallocation of neural substrates. *NeuroImage*, *21*(2), 494–506. https://doi.org/10.1016/j.neuroimage.2003.09.071

Gorina-Careta, N., Kurkela, J. L., Hämäläinen, J., Astikainen, P., & Escera, C. (2021). Neural generators of the frequency-following response elicited to stimuli of low and high frequency: A magnetoencephalographic (meg) study. *NeuroImage*, *231*, 117866. https://doi.org/10.1016/j.neuroimage.2021.117866

Gramfort, A. (2013). Meg and eeg data analysis with mne-python. *Frontiers in Neuroscience*, *7*. https://doi.org/10.3389/fnins.2013.00267

Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., & Garrod, S. (2013). Speech rhythms and multiplexed oscillatory sensory coding in the human brain (D. Poeppel, Ed.). *PLoS Biology*, *11*(12), e1001752. https://doi.org/10.1371/journal.pbio.1001752

Hairston, W. D., Letowski, T. R., & McDowell, K. (2013). Task-related suppression of the brainstem frequency following response (P. V. Nerurkar, Ed.). *PLoS ONE*, *8*(2), e55215. https://doi.org/10.1371/journal.pone.0055215

Hill, K. T., & Miller, L. M. (2009). Auditory attentional control and selection during cocktail party listening. *Cerebral Cortex*, *20*(3), 583–590. https://doi.org/10.1093/cercor/bhp124

Hillyard, S. A., Hink, R. F., Schwent, V. L., & Picton, T. W. (1973). Electrical signs of selective attention in the human brain. *Science*, *182*(4108), 177–180. https://doi.org/10.1126/science.182.4108.177

Holmes, E., Purcell, D. W., Carlyon, R. P., Gockel, H. E., & Johnsrude, I. S. (2017). Attentional modulation of envelope-following responses at lower (93–109 hz) but not higher (217–233 hz) modulation rates. *Journal of the Association for Research in Otolaryngology*, *19*(1), 83–97. https://doi.org/10.1007/s10162-017-0641-9

Holt, L. L., & Lotto, A. J. (2008). Speech perception within an auditory cognitive science framework. *Current Directions in Psychological Science*, *17*(1), 42–46. https://doi.org/10.1111/j.1467-8721.2008.00545.x

Holt, L. L., Tierney, A. T., Guerra, G., Laffere, A., & Dick, F. (2018). Dimension-selective attention as a possible driver of dynamic, context-dependent re-weighting in speech processing. *Hearing Research*, *366*, 50–64. https://doi.org/10.1016/j.heares.2018.06.014

Hoormann, J., Falkenstein, M., & Hohnsbein, J. (1994). Effect of selective attention on the latency of human frequency-following potentials. *NeuroReport*, *5*(13), 1609–1612. https://doi.org/10.1097/00001756-199408150-00017

Hoormann, J., Falkenstein, M., & Hohnsbein, J. (2000). Early attention effects in human auditory-evoked potentials. *Psychophysiology*, *37*(1), 29–42. https://doi.org/10.1111/1469-8986.3710029

Hoormann, J., Falkenstein, M., & Hohnsbein, J. (2004). Effects of spatial attention on the brain stem frequency-following potential. *NeuroReport*, *15*(10), 1539–1542. https://doi.org/10.1097/01.wnr.0000134932.89313.3b

Idemaru, K., & Holt, L. L. (2020). Generalization of dimension-based statistical learning. *Attention, Perception, amp; Psychophysics*, *82*(4), 1744–1762. https://doi.org/10.3758/s13414-019-01956-5

Jaramillo, M. A., & Manos, P. S. (2001). Phylogeny and patterns of floral diversity in the genus piper (piperaceae). *American Journal of Botany*, *88*(4), 706–716. https://doi.org/10.2307/2657072

Jas, M., Engemann, D. A., Bekhti, Y., Raimondo, F., & Gramfort, A. (2017). Autoreject: Automated artifact rejection for meg and eeg data. *NeuroImage*, *159*, 417–429. https://doi.org/10.1016/j.neuroimage.2017.06.030

Jeng, F.-C., Hu, J., Dickman, B., Montgomery-Reagan, K., Tong, M., Wu, G., & Lin, C.-D. (2011). Cross-linguistic comparison of frequency-following responses to voice pitch in american and chinese neonates and adults. *Ear amp; Hearing*, *32*(6), 699–707. https://doi.org/10.1097/aud.0b013e31821cc0df

JENTZSCH, I., & SOMMER, W. (2001). Sequence-sensitive subcomponents of p300: Topographical analyses and dipole source localization. *Psychophysiology*, *38*(4), 607–621. https://doi.org/10.1017/s0048577201000531

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

Kastner, S., & Ungerleider, L. G. (2001). The neural basis of biased competition in human visual cortex. *Neuropsychologia*, *39*(12), 1263–1276. https://doi.org/10.1016/s0028-3932(01)00116-6

Key, A. P. F., Dove, G. O., & Maguire, M. J. (2005). Linking brainwaves to the brain: An erp primer. *Developmental Neuropsychology*, *27*(2), 183–215. https://doi.org/10.1207/s15326942dn2702_1

King, A., Hopkins, K., & Plack, C. J. (2016). Differential group delay of the frequency following response measured vertically and horizontally. *Journal of the Association for Research in Otolaryngology*, *17*(2), 133–143. https://doi.org/10.1007/s10162-016-0556-x

Kramer, S. E., Kapteyn, T. S., Festen, J. M., & Kuik, D. J. (1997). Assessing aspects of auditory handicap by means of pupil dilatation. *International Journal of Audiology*, *36*(3), 155–164. https://doi.org/10.3109/00206099709071969

Kraus, N., Anderson, S., & White-Schwoch, T. (2017). The frequency-following response: A window into human communication. In *The frequency-following response* (pp. 1–15). Springer International Publishing. https://doi.org/10.1007/978-3-319-47944-6_1

Krishnan, A., Xu, Y., Gandour, J., & Cariani, P. (2005). Encoding of pitch in the human brainstem is sensitive to language experience. *Cognitive Brain Research*, *25*(1), 161–168. https://doi.org/10.1016/j.cogbrainres.2005.05.004

Krizman, J., Marian, V., Shook, A., Skoe, E., & Kraus, N. (2012). Subcortical encoding of sound is enhanced in bilinguals and relates to executive function advantages. *Proceedings of the National Academy of Sciences*, *109*(20), 7877–7881. https://doi.org/10.1073/pnas.1201575109

Krizman, J., Skoe, E., Marian, V., & Kraus, N. (2014). Bilingualism increases neural response consistency and attentional control: Evidence for sensory and cognitive coupling. *Brain and Language*, *128*(1), 34–40. https://doi.org/10.1016/j.bandl.2013.11.006

Krizman, J., Slater, J., Skoe, E., Marian, V., & Kraus, N. (2015). Neural processing of speech in children is influenced by extent of bilingual experience. *Neuroscience Letters*, *585*, 48–53. https://doi.org/10.1016/j.neulet.2014.11.011

Lachaux, J.-P., Rodriguez, E., Martinerie, J., & Varela, F. J. (1999). Measuring phase synchrony in brain signals. *Human Brain Mapping*, *8*(4), 194–208. https://doi.org/10.1002/(sici)1097-0193(1999)8:4<194::aid-hbm4>3.0.co;2-c

Lakatos, P., Gross, J., & Thut, G. (2019). A new unifying account of the roles of neuronal entrainment. *Current Biology*, *29*(18), R890–R905. https://doi.org/10.1016/j.cub.2019.07.075

Lange, K. (2009). Brain correlates of early auditory processing are attenuated by expectations for time and pitch. *Brain and Cognition*, *69*(1), 127–137. https://doi.org/10.1016/j.bandc.2008.06.004

Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review*, *106*(1), 119–159. https://doi.org/10.1037/0033-295x.106.1.119

Lehmann, A., & Schönwiesner, M. (2014). Selective attention modulates human auditory brainstem responses: Relative contributions of frequency and spatial cues (J. Ahveninen, Ed.). *PLoS ONE*, *9*(1), e85442. https://doi.org/10.1371/journal.pone.0085442

Lenth, R. V. (2025). Emmeans: Estimated marginal means, aka least-squares means [R package version 1.11.2]. https://CRAN.R-project.org/package=emmeans

Lerud, K. D., Hancock, R., & Skoe, E. (2023). A high-density eeg and structural mri source analysis of the frequency following response to missing fundamental stimuli reveals subcortical and cortical activation to low and high frequency stimuli. *NeuroImage*, *279*, 120330. https://doi.org/10.1016/j.neuroimage.2023.120330

Li, L., Miller, E. K., & Desimone, R. (1993). The representation of stimulus familiarity in anterior inferior temporal cortex. *Journal of Neurophysiology*, *69*(6), 1918–1929. https://doi.org/10.1152/jn.1993.69.6.1918

Li, Z. (2002). Pop-out theory: Segmentation without classification by the primary visual cortex. In *Visual attention mechanisms* (pp. 69–78). Springer US. https://doi.org/10.1007/978-1-4615-0111-4_7

Luck, S. J., & Hillyard, S. A. (1994). Spatial filtering during visual search: Evidence from human electrophysiology. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(5), 1000–1014. https://doi.org/10.1037/0096-1523.20.5.1000

Luck, S. J., & Kappenman, E. S. (2011, December). *Erp components and selective attention*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195374148.013.0144

Luo, H., & Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron*, *54*(6), 1001–1010. https://doi.org/10.1016/j.neuron.2007.06.004

McAdams, C. J., & Maunsell, J. H. R. (1999). Effects of attention on orientation-tuning functions of single neurons in macaque cortical area v4. *The Journal of Neuroscience*, *19*(1), 431–441. https://doi.org/10.1523/jneurosci.19-01-00431.1999

Mireille, B. (2010). The musician brain as an excellent model of brain plasticity. *Frontiers in Neuroscience*, *4*. https://doi.org/10.3389/conf.fnins.2010.12.00038

Musacchia, G., Strait, D., & Kraus, N. (2008). Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and non-musicians. *Hearing Research*, *241*(1–2), 34–42. https://doi.org/10.1016/j.heares.2008.04.013

Näätanen, R., Teder, W., Alho, K., & Lavikainen, J. (1992). Auditory attention and selective input modulation: A topographical erp study. *NeuroReport*, *3*(6), 493–496. https://doi.org/10.1097/00001756-199206000-00009

Näätänen, R., Gaillard, A., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica*, *42*(4), 313–329. https://doi.org/10.1016/0001-6918(78)90006-9

Näätänen, R. (1990). The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. *Behavioral and Brain Sciences*, *13*(2), 201–233. https://doi.org/10.1017/s0140525x00078407

Näätänen, R., & Winkler, I. (1999). The concept of auditory stimulus representation in cognitive neuroscience. *Psychological Bulletin*, *125*(6), 826–859. https://doi.org/10.1037/0033-2909.125.6.826

Nieuwenhuis, S., Yeung, N., van den Wildenberg, W., & Ridderinkhof, K. R. (2003). Electrophysiological correlates of anterior cingulate function in a go/no-go task: Effects of response conflict and trial type frequency. *Cognitive, Affective, amp; Behavioral Neuroscience*, *3*(1), 17–26. https://doi.org/10.3758/cabn.3.1.17

Obleser, J., & Kayser, C. (2019). Neural entrainment and attentional selection in the listening brain. *Trends in Cognitive Sciences*, *23*(11), 913–926. https://doi.org/10.1016/j.tics.2019.08.004

O'Craven, K. M., Downing, P. E., & Kanwisher, N. (1999). Fmri evidence for objects as the units of attentional selection. *Nature*, *401*(6753), 584–587. https://doi.org/10.1038/44134

Oxenham, A. J. (2013). Revisiting place and temporal theories of pitch. *Acoustical science and technology*, *34*(6), 388–396.

Paltoglou, A. E., Sumner, C. J., & Hall, D. A. (2009). Examining the role of frequency specificity in the enhancement and suppression of human cortical activity by auditory selective attention. *Hearing Research*, *257*(1–2), 106–118. https://doi.org/10.1016/j.heares.2009.08.007

Pantev, C., Oostenveld, R., Engelien, A., Ross, B., Roberts, L. E., & Hoke, M. (1998). Increased auditory cortical representation in musicians. *Nature*, *392*(6678), 811–814. https://doi.org/10.1038/33918

Parbery-Clark, A., Skoe, E., & Kraus, N. (2009). Musical experience limits the degradative effects of background noise on the neural processing of sound. *The Journal of Neuroscience*, *29*(45), 14100–14107. https://doi.org/10.1523/jneurosci.3256-09.2009

Patel, G. H., Snyder, L. H., & Corbetta, M. (2018). Attentional modulation of macaque visual processing areas. https://doi.org/10.1101/345439

Peelle, J. E. (2018). Listening effort: How the cognitive consequences of acoustic challenge are reflected in brain and behavior. *Ear amp; Hearing*, *39*(2), 204–214. https://doi.org/10.1097/aud.0000000000000494

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). Psychopy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Phillips, N. A. (2016). The implications of cognitive aging for listening and the framework for understanding effortful listening (fuel). *Ear amp; Hearing*, *37*(1), 44S–51S. https://doi.org/10.1097/aud.0000000000000309

Polich, J. (2007). Updating p300: An integrative theory of p3a and p3b. *Clinical Neurophysiology*, *118*(10), 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019

Price, C. N., & Bidelman, G. M. (2020). Attention reinforces human corticofugal system to aid speech perception in noise. https://doi.org/10.1101/2020.10.22.351494

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87. https://doi.org/10.1038/4580

Reynolds, J. H., Chelazzi, L., & Desimone, R. (1999). Competitive mechanisms subserve attention in macaque areas v2 and v4. *The Journal of Neuroscience*, *19*(5), 1736–1753. https://doi.org/10.1523/jneurosci.19-05-01736.1999

Rönnberg, J., Rudner, M., Foo, C., & Lunner, T. (2008). Cognition counts: A working memory system for ease of language understanding (elu). *International Journal of Audiology*, *47*(sup2), S99–S105. https://doi.org/10.1080/14992020802301167

Rosenthal, F. (1963). Extension of single-factor nerve excitation theory to cortical responses. *Experimental Neurology*, *8*(2), 125–131. https://doi.org/10.1016/0014-4886(63)90039-6

Russo, N. M., Nicol, T. G., Zecker, S. G., Hayes, E. A., & Kraus, N. (2005). Auditory training improves neural timing in the human brainstem. *Behavioural Brain Research*, *156*(1), 95–103. https://doi.org/10.1016/j.bbr.2004.05.012

Schaffer, A. (2015, August). Faculty opinions recommendation of microbial species delineation using whole genome sequences. https://doi.org/10.3410/f.725620247.793508999

SCHLAUG, G. (2001). The brain of musicians: A model for functional and structural adaptation. *Annals of the New York Academy of Sciences*, *930*(1), 281–299. https://doi.org/10.1111/j.1749-6632.2001.tb05739.x

Schroeder, C. E., & Lakatos, P. (2009). Low-frequency neuronal oscillations as instruments of sensory selection. *Trends in Neurosciences*, *32*(1), 9–18. https://doi.org/10.1016/j.tins.2008.09.012

Schüller, A., Schilling, A., Krauss, P., Rampp, S., & Reichenbach, T. (2023). Attentional modulation of the cortical contribution to the frequency-following response evoked by continuous speech. https://doi.org/10.1101/2023.07.03.547608

Schwartz, Z. P., & David, S. V. (2017). Focal suppression of distractor sounds by selective attention in auditory cortex. *Cerebral Cortex*, *28*(1), 323–339. https://doi.org/10.1093/cercor/bhx288

Shamma, S., & Fritz, J. (2014). Adaptive auditory computations. *Current Opinion in Neurobiology*, *25*, 164–168. https://doi.org/10.1016/j.conb.2014.01.011

Skoe, E., Burakiewicz, E., Figueiredo, M., & Hardin, M. (2017). Basic neural processing of sound in adults is influenced by bilingual experience. *Neuroscience*, *349*, 278–290. https://doi.org/10.1016/j.neuroscience.2017.02.049

Skoe, E., & Kraus, N. (2010). Auditory brain stem response to complex sounds: A tutorial. *Ear amp; Hearing*, *31*(3), 302–324. https://doi.org/10.1097/aud.0b013e3181cdb272

Snyder, J. S., & Alain, C. (2007). Toward a neurophysiological theory of auditory stream segregation. *Psychological Bulletin*, *133*(5), 780–799. https://doi.org/10.1037/0033-2909.133.5.780

Sommer, W., Leuthold, H., & Soetens, E. (1999). Covert signs of expectancy in serial reaction time tasks revealed by event-related potentials. *Perception amp; Psychophysics*, *61*(2), 342–353. https://doi.org/10.3758/bf03206892

Song, J. H., Skoe, E., Wong, P. C. M., & Kraus, N. (2008). Plasticity in the adult human auditory brainstem following short-term linguistic training. *Journal of Cognitive Neuroscience*, *20*(10), 1892–1902. https://doi.org/10.1162/jocn.2008.20131

Strait, D. L., O'Connell, S., Parbery-Clark, A., & Kraus, N. (2013). Musicians' enhanced neural differentiation of speech sounds arises early in life: Developmental evidence from ages 3 to 30. *Cerebral Cortex*, *24*(9), 2512–2521. https://doi.org/10.1093/cercor/bht103

Strait, D. L., Kraus, N., Skoe, E., & Ashley, R. (2009). Musical experience and neural efficiency – effects of training on subcortical processing of vocal expressions of emotion. *European Journal of Neuroscience*, *29*(3), 661–668. https://doi.org/10.1111/j.1460-9568.2009.06617.x

Suga, N., Xiao, Z., Ma, X., & Ji, W. (2002). Plasticity and corticofugal modulation for hearing in adult animals. *Neuron*, *36*(1), 9–18. https://doi.org/10.1016/s0896-6273(02)00933-9

Sutton, S., Braren, M., Zubin, J., & John, E. R. (1965). Evoked-potential correlates of stimulus uncertainty. *Science*, *150*(3700), 1187–1188. https://doi.org/10.1126/science.150.3700.1187

Taylor, M. J. (1996). Development of early visual selective attention processes studied with event-related brain potentials: (526132012-064). https://doi.org/10.1037/e526132012-064

Theunissen, F. (2003, December). Faculty opinions recommendation of rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. https://doi.org/10.3410/f.1016670.201144

Tichko, P., & Skoe, E. (2017). Frequency-dependent fine structure in the frequency-following response: The byproduct of multiple generators. *Hearing Research*, *348*, 1–15. https://doi.org/10.1016/j.heares.2017.01.014

Tierney, A., Krizman, J., Skoe, E., Johnston, K., & Kraus, N. (2013). High school music classes enhance the neural processing of speech. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00855

Tierney, A. T., Krizman, J., & Kraus, N. (2015). Music training alters the course of adolescent auditory development. *Proceedings of the National Academy of Sciences*, *112*(32), 10062–10067. https://doi.org/10.1073/pnas.1505114112

Todorovic, A., van Ede, F., Maris, E., & de Lange, F. P. (2011). Prior expectation mediates neural adaptation to repeated sounds in the auditory cortex: An meg study. *Journal of Neuroscience*, *31*(25), 9118–9123. https://doi.org/10.1523/jneurosci.1425-11.2011

Treue, S., & Trujillo, J. C. M. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature*, *399*(6736), 575–579. https://doi.org/10.1038/21176

Turatto, M., Mazza, V., & Umiltà, C. (2005). Crossmodal object-based attention: Auditory objects affect visual processing. *Cognition*, *96*(2), B55–B64. https://doi.org/10.1016/j.cognition.2004.12.001

Tzounopoulos, T., & Kraus, N. (2009). Learning to encode timing: Mechanisms of plasticity in the auditory brainstem. *Neuron*, *62*(4), 463–469. https://doi.org/10.1016/j.neuron.2009.05.002

Volosin, M., & Horváth, J. (2020). Task difficulty modulates voluntary attention allocation, but not distraction in an auditory distraction paradigm. *Brain Research*, *1727*, 146565. https://doi.org/10.1016/j.brainres.2019.146565

Wang, X., Lu, T., Bendor, D., & Bartlett, E. (2008). Neural coding of temporal information in auditory thalamus and cortex. *Neuroscience*, *157*(2), 484–493. https://doi.org/10.1016/j.neuroscience.2008.07.050

Winn, M. B., Edwards, J. R., & Litovsky, R. Y. (2015). The impact of auditory spectral resolution on listening effort revealed by pupil dilation. *Ear amp; Hearing*, *36*(4), e153–e165. https://doi.org/10.1097/aud.0000000000000145

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

Wong, P. C. M., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience*, *10*(4), 420–422. https://doi.org/10.1038/nn1872

Woods, D. L., Alain, C., Diaz, R., Rhodes, D., & Ogawa, K. H. (2001). Location and frequency cues in auditory selective attention. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(1), 65–74. https://doi.org/10.1037/0096-1523.27.1.65

Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: Music and speech. *Trends in Cognitive Sciences*, *6*(1), 37–46. https://doi.org/10.1016/s1364-6613(00)01816-7

Zatorre, R. J., Bouffard, M., & Belin, P. (2004). Sensitivity to auditory object features in human temporal neocortex. *The Journal of Neuroscience*, *24*(14), 3637–3642. https://doi.org/10.1523/jneurosci.5458-03.2004

Zatorre, R. J., Delhommeau, K., & Zarate, J. M. (2012). Modulation of auditory cortex response to pitch variation following training with microtonal melodies. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00544

Zhang, Y., Lehmann, A., & Deroche, M. (2020). Disentangling listening effort and memory load beyond behavioural evidence: Pupillary response to listening effort during a concurrent memory task. https://doi.org/10.1101/2020.05.04.076588

Zhu, L., Bharadwaj, H., Xia, J., & Shinn-Cunningham, B. (2013). A comparison of spectral magnitude and phase-locking value analyses of the frequency-following response to complex tones. *The Journal of the Acoustical Society of America*, *134*(1), 384–395. https://doi.org/10.1121/1.4807498

Additional Materials

Appendix A: Music and background questionnaire

Background questionnaire

* Indicates required question

1. Subject number *

____

Demographics

Below are questions about your education and language use. Please answer them as completely as possible.

2. Date of birth *

____

Example: January 7, 2019

3. Country or countries of origin *

____

4. Undergraduate major *

____

5. Highest degree of education obtained * ⓥ Dropdown

Mark only one oval.

○ High school diploma
○ Bachelor's
○ Master's
○ PhD
○ Other

6. Gender Identity (select all that apply): *

Check all that apply.

☐ Woman
☐ Man
☐ Transgender
☐ Non-binary/non-conforming
☐ Prefer not to respond
☐ Other: ______________________________

Language background

7. Is English your native language? *

Mark only one oval.

○ Yes
○ No

8. At what age did you start learning English *

____

9. Please list any other languages you are a native speaker of *

____

10. Please list any other languages you have learned and how many years you have learned them for (e.g. Spanish, 5 years) *

Speech or hearing impairments

11. Do you have any known speech or hearing impairments? *

Music experience

12. Do you or have you ever played any musical instruments? *

Mark only one oval.

○ Yes
○ No

13. Do you still actively play any musical instruments? *

Mark only one oval.

○ Yes
○ No

14. If you have played an instrument at any point, at what at what age did you first begin musical instruction?

15. If you have played an instrument at any point, at what age did you stop having these lessons?

16. If you have played an instrument at any point, please list all the musical instruments you have played and the number of years you have taken lessons for them (e.g. violin, 2 years; music theory, 1 year)

17. If you have played an instrument at any point, please list all the musical instruments you have played and the number of years you have actively played them for (e.g. violin, 5 years; bass, 1 year)

18. Have you had any type of vocal training, including choir or a capella? *

Mark only one oval.

○ Yes
○ No

19. If yes, how many years of lessons have you had? *

____

20. Have you taken dance lessons? *

Mark only one oval.

○ Yes
○ No

21. If yes, how many years of lessons have you had?

____

22. Approximately how many hours do you spend making music per week? *

____

23. Approximately how many hours do you spend listening to music per day? *

____

24. Please add any extra information you think is necessary

____
____
____
____
____