THE UNIVERSITY OF CHICAGO

OPERATION AND INTERACTION OF THE DESCENDING CONTROL AND ASCENDING SENSORY PATHWAYS IN AUDITION

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
SERENA KLOS

CHICAGO, ILLINOIS
JUNE 2016

This dissertation is dedicated to my grandparents, Howard and Eleanor Klos.

Table of Contents

List of Tables

List of Figures

Acknowledgements

The completion of this dissertation would not have been possible without the help and support of many important and wonderful individuals. To start, I would like to thank my advisor for his unwavering encouragement and guidance throughout all of these years. Howard Nusbaum has been an incredible source of support and knowledge since the first day of my graduate career. His kindness seems to know no bounds, nor does his awe-inspiring familiarity with practically every realm of Psychology. Most importantly, he gave me the freedom, resources and confidence to pursue the research that truly interested and inspired me. Howard, thank you for always believing in my capabilities and for reminding me of my accomplishments when it was difficult for me to see past the setbacks.

I would also like to thank my dissertation committee. Dan Margoliash provided unwavering support and invaluable input during my pursuit of mastering the art of the auditory brainstem response. Leslie Kay had a tremendous impact not only on the way in which I planned and interpreted these experiments, but also on how I think about research in general, which will never leave me. Finally, Marc Berman gave me valuable insight into how I could situate my research within a broader context and wider areas of research.

This research would not have been possible without the help and volunteerism of many past and present members of the Nusbaum Lab. To every lab member who volunteered his/her head (and ears) to conducting a quality control test for the equipment itself or a pilot test for my own experiments, I am incredibly grateful for your time and patience as I tortured you by putting electrodes on your scalps and then forcing you to listen to the same annoying sound thousands of times. In particular, I must thank George Abitante, who underwent more equipment and pilot tests than anyone else and still showed up each day with a smile on his face. I must also thank

Mike Pardo, for without his willingness to spend countless hours clicking buttons to complete data preprocessing, this data would never have been analyzed in time. I would also like to thank Mike for his frequent words of encouragement and positive assessments of the situation, particularly when I was unable to do so myself. Shannon Heald and Steve Van Hedger – two thirds of the dynamic Triple S, Triple Threat. Words cannot describe your impact on my academic and non-academic life. Thank you for wonderful scientific discussions that always challenged me to be a better scientist and much-needed non-scientific conversations that kept me sane.

I would also like to extend my gratitude to my family in New York, who has shown unconditional support for my decision to pursue a doctoral degree. Each member of the Klos family, Mom, Grandma, Rachael, Marisa, Nick and Cole, has supported me in their own unique way. Of course, I would be remiss if I did not thank my Ace's family for all of their support as well. Joe and Harry, thank you for providing a place for me to escape academia and recharge for a bit while also giving me invaluable life advice.

Finally, there are two individuals without whom I would not have survived my UChicago experience. Erin Cable taught me the value of a good super hero movie, was always willing to go on an adventure to see what this city had to offer, and enthusiastically partook in any dish that I attempted to cook, thus fueling my love of cooking and my need to share it with others. Most importantly, she encouraged me to adopt my furry companion Lily, who has provided me with much needed comic relief, reminded me that it is important to be outside as much as possible and given me countless hours of happiness. And of course, thanks to Emma Von Jetson for always being my biggest fan.

Abstract

Neural theories of human auditory perception often oversimplify the interaction of afferent and efferent projections in the corticofugal pathway. However, investigations of the effects of long-term experience on the auditory brainstem response demonstrate that cortical systems can affect early processing of acoustic information through top-down efferent projections (Krishnan et al., 2005; Wong et al., 2007). The immediate or real time effects of these top-down pathways on activity in the auditory brainstem, however, are less clear (Galbraith et al., 1998; Hoormann et al., 1994). One approach to investigating possible corticofugal interactions is to examine whether the suppressive effects of motor behavior exhibited in primary auditory cortex (Houde et al., 2002; Schneider et al., 2014) can be demonstrated at the level of the auditory brainstem and to determine whether these effects are driven by divided attention between the motor and auditory systems. Chapters 2 and 3 tested for motor behavior versus attention effects on neural activity in the auditory brainstem resulting from the presentation of a simple tone or a brief synthetic syllable, with the hypothesis that motor behavior or separable attention demands could reduce the auditory brainstem response to the acoustic stimulus. Chapter 4 held motor behavior constant and varied the demands on attention. In all cases, attention to moments in time enhanced spectral responses in the auditory brainstem response to acoustic stimuli, rejecting the hypothesis that there is descending cortical inhibition of auditory responses in the brainstem due to motor behavior, and the hypothesis that there are descending effects from the cortex of dividing attention. The results suggest that a system-wide attention network that directs attention to specific events or moments in time exerts control over the descending auditory pathway. Importantly, the results demonstrate the need to include the

bidirectional projections between cortical networks and subcortical sensory structures in neural accounts of human auditory and speech perception.

Chapter 1: Introduction

Human sensory systems, such as vision and audition, are generally understood as transducing physical energy into neural responses that are transmitted from sensory receptors to cortical systems for processing (Wolfe et al., 2012). Brainstem areas in the early sensory pathway are often viewed as transforming the neural patterns of encoded stimuli into a form to be organized processed and ultimately recognized by cortical systems (Recanzone & Sutter, 2008). This essentially conceives of sensory information processing as a predominantly bottom-up system. The idea of top-down processing, perceptual mechanisms that direct sensory information processing based on knowledge or expectations, is often limited to the modification or interaction with primary visual or auditory cortex, despite the fact that descending inputs to the cochlea are well known (Moore, 1987). When subcortical interactions are posited, they are described as neuronal reorganization processes as in the corticofugal system in the bat (Suga et al., 2002) or the thalamocortical interactions in visual attention (Spinelli & Pribram, 1966; although see Sherman, 2006; 2009). Even when descending efferent connections within human sensory systems are identified, as has long been the case in the auditory system (Rasmussen, 1946; Schwartz, 1992; Moore, 1987) or more recently in vision (Akimov et al., 2010; Gastinger et al., 2006; Resch et al., 2005), there is scant evidence that these descending projections are used to modify peripheral sensory responses.

Throughout auditory system research, there is a deep discrepancy between the knowledge that, across multiple species, for every afferent projection from the cochlear nucleus to primary auditory cortex, there is also a parallel efferent projection (Huffman & Henson, 1990), evidence from animal models demonstrating that cortical structures utilize the corticofugal system to engage in egocentric selection (Suga et al., 2002) and current models of human auditory and speech perception. Research on the descending auditory system in humans has largely focused on subcortical interactions that act as part of the noise protection circuit that prevents damage to the inner ear (Brown, 2011), rather than as part of the control system that produces sensitive hearing. Similarly, accounts of speech perception highlight top-down cortico-cortical interactions primarily between motor, attention and auditory networks (Rauschecker & Scott, 2009; Skipper, 2014), without acknowledging the growing evidence for top-down cortical effects on the auditory brainstem and the cochlea (Galbraith & Arroyo, 1993; Giard et al., 1994; Maison et al., 2001). This dissertation is focused on investigating the possibility that these descending neural projections, both from the auditory brainstem to inner and outer hair cells and from cortical networks to the peripheral auditory pathway, represent part of a control system for modifying auditory responses as a result of top-down attention. It will outline the current knowledge of the auditory circuit from the cochlea to cortex and demonstrations in animal models of the influence of the corticofugal system on early auditory perception. The dissertation will then discuss the discrepancies between human models of auditory perception and models of speech perception, given the evidence for top-down attentional effects on the peripheral auditory pathway, before proposing a perceptual system controlled from the top down, based on contextually determined expectations. Such a perceptual system confers an advantage in perception of flexibility to meet the challenges of adverse and dynamically changing sensory environments (e.g. varying high noise levels) as well as increasing recognition for ambiguous patterns by shifting the basis for classification.

Acoustic input is transmitted across several structures along the auditory pathway before reaching the primary auditory cortex. As illustrated in Figure 1.1, the auditory signal is first transferred from each cochlea to the cochlear nucleus. This is the only set of projections that remain completely ipsilateral; from this point on in the auditory pathway, acoustic input from each ear travels ipsliaterally and contralaterally to primary auditory cortex. Ascending fibers are then projected from the cochlear nucleus to the superior olivary complex and lateral lemniscus. The superior olivary complex also has ascending projections to the lateral lemniscus and sends further fibers to both inferior colliculi (Picton et al., 1981). Input is then transmitted from the inferior colliculi to both medial geniculate bodies in the thalamus, which ultimately send ascending projections to the primary auditory cortex (Chandrasekaran & Kraus, 2010). Figure 1.1 also exemplifies an intricate system of descending fibers that affect every level of the peripheral auditory pathway. According to Huffman and Henson (1990), this descending pathway can be divided into three major projections: the auditory cortex to the medial geniculate body and the inferior colliculus, the inferior colliculus to the cochlear nucleus and the superior olivary complex, and the superior olivary complex to the cochlea. The authors attribute the first two major projections as a feedback system that promotes a tonic increase in excitability in lower auditory structures combined with a stimulus-evoked increase in excitability in the auditory cortex to create an enhanced and faster evoked response in the auditory cortex. The third efferent projection, between the superior olivary complex and the cochlea, has been well established as a set of medial and lateral olivocochlear efferents which directly synapse on outer and inner hair cells, respectively, and can inhibit basilar membrane responses to both reduce acoustic trauma and reduce the masking of speech by background noise (Guinan, 2006). Chandrasekaran and Kraus (2010) argue for further descending projections from the auditory cortex to the superior olivary complex and the cochlear nucleus as well to demonstrate the effects of cortical network activity on the representation of acoustic information in auditory brainstem structures.

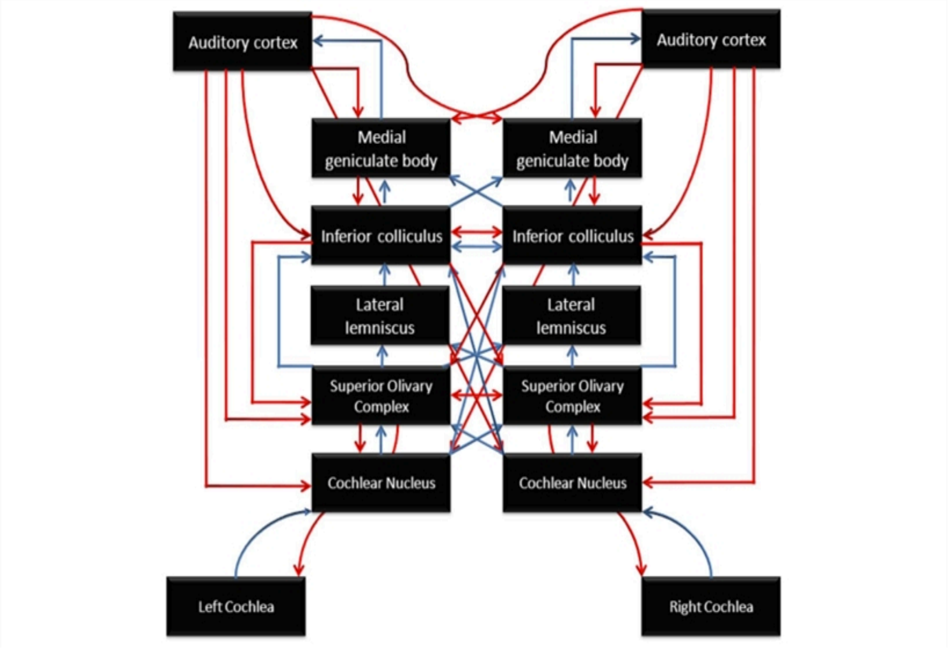

Figure 1.1. Schematic illustration of the auditory system from Chandrasekaran & Kraus, 2010. Blue arrows correspond to the ascending (bottom-up) pathways; red arrows correspond to the descending projections.

Animal models of the auditory system often refer to the corticofugal pathway as a neuronal reorganization system in which cortical structures guide the reconfiguration of the receptive fields of lower structures. Suga and colleagues (2002), for example, demonstrated that cortical structures use the corticofugal system to engage in egocentric selection, in which they improve their own input from the brainstem through feedback and lateral inhibition. Such processes allow for rapid readjustment of subcortical processing and long-term adjustments in cortex to facilitate associative learning. Others have shown in the monkey visual system that efferent projections from the inferotemporal cortex alter the configuration of receptive fields in the thalamus to promote selective visual attention (Spinelli & Pribram, 1966), suggesting that most corticofugal models focus mainly on the modality specific activities that occur within the these descending projections to promote the encoding of certain sensory inputs over others. Sherman (2006; 2009), however, proposes a dynamic interaction between the thalamus and cortex in particular, with efferent corticothalamic projections acting as mediators between the cortex and other subcortical structures as well as an active partner in cortical communications.

Examinations of the corticofugal system in humans have often been limited to the activation of a noise protection circuit that prevents damage to the inner ear (Brown, 2011), despite the fact that descending projections to the inner hair cells could modify the receptive field properties of the ear (Adams, 1995; Thompson & Thompson, 1991) and descending projections to the outer hair cells do change the relative stiffness of the basilar membrane response to sound (Cooper & Guinan, 2006) thus providing a mechanism to further change frequency or pattern sensitivity in the cochlea. Research has shown that this system does seem to be affected in a noisy environment (Dolan & Nutall, 1988; Kawase, Delgutte, & Liberman, 1993; Winslow & Sachs, 1988) supporting the notion of a noise protection circuit, but research has also demonstrated that higher-level cognitive functions can have an effect on structures in the early sensory pathway as well. Selective attention has been shown to enhance the spectral peaks of evoked otoacoustic emissions (Giard et al., 1994; Maison et al., 2001) and discrimination training has been directly related to enhanced suppression of click-evoked otoacoustic emissions (de Boer & Thornton, 2008). The growing evidence that the human corticofugal system plays an important role in audition suggests not only that the influence of top-down cortical projections may be seen throughout the peripheral auditory nervous system but also that these descending pathways should be included in any neural model of human perception in order to accurately portray the flexibility necessary within the system to adapt to constantly changing sensory environments.

Neural accounts of human auditory perception, particularly in the realm of speech comprehension, have attempted to describe a mechanism for perceptual flexibility at various junctures of the auditory pathway. Abbs and Sussman (1971), drawing from the organizational network of feature specific neurons in the visual system, suggested that the mapping of complex acoustic signals onto phonetic categories could be achieved with a similar set of neurons specialized to identify certain spectrotemporal features of the acoustic signal, such as formant frequency, intensity and rate change. The perception and transmission of certain acoustic features over others could be achieved through a form of lateral inhibition in the peripheral auditory pathway (as low as the basilar membrane) in which active neurons would inhibit adjacent receptors specific to other acoustic features so as to sharpen the detection of input at the lowest levels of perception. This description of auditory perception allows for a contextually based neural tuning to specific acoustic properties of the incoming signal at the earliest levels of sensation and perception, but it does not explain how certain features are deemed more important or salient than others. Abbs and Sussman do not discuss the basis for the lateral inhibition that may occur, making it unclear what cortical networks, if any, may be driving the selection of specific acoustic features at the level of the peripheral auditory system.

More recent neural models of auditory perception have been derived from dual pathway neural networks described in visual perception. These models were devised to explain the dissociation that appears to exist both functionally and anatomically in vision between identifying where an object is located in space and as what the object can be identified. They consist of two different pathways, a dorsal stream and a ventral steam, each of which is responsible for one of these two functions (Ungerleider & Mishkin, 1982; Figure 1.2). While the dorsal stream, the purported seat of object localization (Milner & Goodale, 1995) is typically described as completely separate from the object categorizing ventral stream, Bar (2003) proposed that the dorsal stream may project through prefrontal cortex and ultimately connect with the ventral pathway in order to facilitate coarse object classification recognition. This characterization of the dorsal and ventral streams implies that active feedforward and feedback processes take place both within the visual perception network as well as with other cognitive networks involved in attention, working memory and executive function in order to achieve object perception. Several models of speech perception have utilized this neurobiological architectural distinction of dorsal and ventral streams to varying degrees. While some have adhered to the functional and structural separation of the dorsal and ventral streams, others rely on the interaction of the two pathways to create feedforward and feedback pathways within the auditory network.

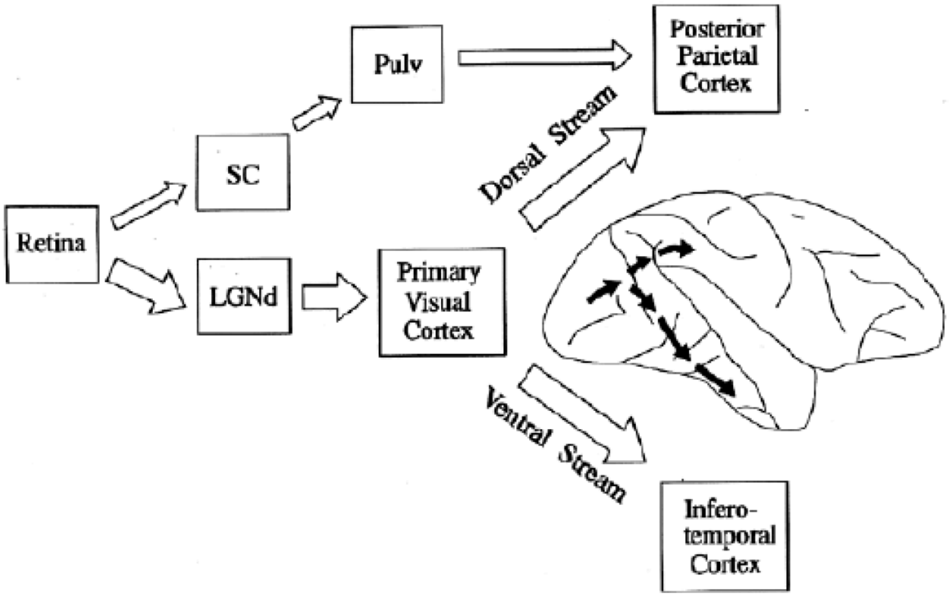

Figure 1.2. The major routes of visual input into the dorsal and ventral streams. From Milner and Goodale (1995).

Unlike the models proposed in vision, however, all of these models have largely adhered to the cortical structures of the perisylvian language network, without fully venturing into other neural networks that may play an important role in the recognition of auditory objects. For example, some neural speech models have relied on suppositions of general auditory theory to describe auditory perception. The general auditory hypothesis postulates that a speaker utilizes his/her knowledge of the acoustic patterns of a vocalization to elicit the appropriate phonetic categories in the listener (Diehl, Kluender, Walsh and Parker, 1991). This speech recognition process employs the same mechanisms implemented for general auditory processing in humans, meaning that statistical regularities in the distribution of stimulus properties within the speech signal are used to create speech categories in the same way that general perceptual categories are developed in other forms of sensory processing (Holt and Lotto, 2008). Drawing from this cognitive model, Hickok and Poeppel (2007) have proposed a neurobiological network that explicitly separates ventral and dorsal speech processing streams, identifying these largely with speech object recognition (ventral) and speech perception-production (dorsal). There is a somewhat unusual distinction made in this theory between speech perception and speech recognition as processes that double dissociate both functionally and cortically. Hickok and Poeppel define speech perception (as dissociated from speech recognition) as any sublexical task that involves the discrimination or categorization of auditory input. It is an active process that requires both working memory and executive control, but does not necessarily lead to the lexical-sentential understanding of the speech signal. Such a description of the dorsal stream might imply that speech perception is reliant upon processes occurring in other cognitive networks to accurately identify auditory objects. However this is not how Hickok and Poeppel, at least in their initial description of the model, describe the dorsal stream which seems more functionally focused on word learning and metalinguistic task performance in speech perception experiments, but is not identified as having any role in recognizing spoken language, although the connections are present within the model for this possibility. By contrast, utterances are recognized and understood by the process of speech recognition, which takes place solely within the ventral stream, transforming acoustic signals into mental lexicon representations. In subsequent descriptions of the model Hickok and colleagues (2011) have included interactions between the ventral and dorsal pathways in instances of particularly difficult or noisy speech input, but they do not describe a system in which there is a constant forward and inverse interaction between the two auditory streams based on changes in attention and executive control functions, as has been emphasized in visual neural models.

The exclusions of additional cognitive networks in this and similar neural models of speech perception may be driven by the continued debate as to the degree to which perception is an active process. In general, when conceiving of a system that can map acoustic input into a percept, two broad classes of computational theories may be envisioned: passive processes and active processes (Nusbaum & Schwab, 1986). The distinction between these two systems refers to the manner in which a stimulus is transformed and analyzed within a set of neural networks until perception occurs. In a passive process, sensory input is transmitted and matched to a perceptual representation in an invariant and sequential manner so that the acoustic properties of a signal are interpreted as the same categories each time that they are encountered. Of course, this is probably not the case for sensory processing in waking animals and is certainly not the case for spoken language. Speech is characterized by a many-to-many mapping, meaning that not only can many different acoustic forms represent the same linguistic category, in a many-to-one format, but the same acoustic input may be interpreted as different percepts in each instance that it is produced (a one-to-many format) (Nusbaum & Magnuson, 1997). This creates a lack of invariance problem for speech perception in which variance is constantly present in all aspects of the acoustic to phonetic mapping.

The model proposed by Hickok and Poeppel most closely adheres to a passive approach to speech perception, as it separates the dorsal and ventral streams both functionally and structurally and only rarely accounts for an interaction between the two that may be facilitated by contextual demands. The model includes many of the pieces necessary to create a flexible perceptual system that incorporates inputs from other cortical networks, but the lack of feedforward and feedback interactions between the two processing streams makes it impossible to do so. It cannot account for the recognition of new or altered acoustic combinations that may not match with the acoustic patterns of vocalizations already stored in memory nor can it explain how changes in the internal state of the listener, such as a redirection of focused attention or changes in expectations, can influence recognition, as it limits the neural architecture of speech perception to only those regions dedicated to audition. Most importantly, for the purposes of this set of experiments, the model ignores any interactions that may occur between cortical networks and subcortical structures to modify the transmission of acoustic signal in the peripheral auditory pathway

An active computational system is more successful in accounting for the lack of invariance problem encountered in perception. An active process is a computational system in which the specific analyses of a signal are contingent upon previously carried out computations or transformations. Active processes can take on either a bottom-up or a top-down driven form. In a bottom-up, data driven system input is still initially processed in a passive manner along subcortical structures so that it is transformed into an initial representation, but it is then susceptible to feedback signals from higher cortical structures that process contextual input. If this contextually based feedback does not match the initial representation of the input, a new representation of the signal will be generated at a subcortical level of processing (Heald &

Nusbaum, 2014). Conversely, a top-down driven active process neurally codes input in a fairly passive manner along the subcortical pathway until it is compared with an internally generated, knowledge based hypothesis or expectation about the input somewhere within cortex (although see models of active sensing, e.g. Morillon et al., 2015). If a mismatch between the expectation and the input occurs, a new hypothesis is generated and the process is repeated until a match is reached between the two and an acceptable percept has been identified (Heald & Nusbaum, 2014). This form of active processing is a conceptually based system in which expectations about the signal are generated in a top-down manner from a knowledge base. Both of these types of active processes describe perception as a form of hypothesis testing that can quickly adapt to changes in the acoustic environment through the use of selective attention and working memory. While they require the use of these cognitive networks, they do not limit the nature of the knowledge that could be used in recognition. Evidence suggests that both forms of active processes can explain at least some of the flexibility evident in auditory and speech perception, however the experiments presented in this dissertation demonstrate that early auditory processes are most influenced by top-down projections that are driven by expectations generated from contextual factors.

Several neurobiological accounts (e.g. Davis and Johnsrude, 2007; Friederici, 2012; Rauschecker & Scott, 2009) have described speech perception as an active process to varying degrees. These models have relied on the networks proposed by motor theory to create a more direct interaction between different pathways and brain regions. Motor theory proposes that knowledge of speech production reduces the uncertainty of mapping acoustic properties onto linguistic categories in perception (Liberman et al., 1967). Rather than grouping speech with other forms of auditory input, motor theorists posit that speech is a special code that requires a specific decoder in order to be recognized. This decoder is found in the overlapping articulators that are used to produce speech. When speech is heard, these same articulators are utilized to decode the set of motor commands that were used by the speaker to produce those sounds.

The neural accounts most closely associated with motor theory all include the same distinctive interaction between certain structures within the temporal, frontal and parietal cortices. In general, these neural systems include some form of a ventral pathway from primary auditory cortex and the anterior portion of the superior temporal gyrus (STG) to the inferior frontal gyrus (IFG), which is responsible for an initial decoding of the acoustic signal into linguistic categories, as well as a dorsal pathway that originates in the posterior portion of STG and extends through various regions of inferior parietal cortex (IPC) to the premotor cortex and utilizes knowledge of the movements performed by the articulators during speech production to modify or adjust the acoustic to phonetic mapping that has been performed in the ventral pathway through both feedback and feedforward mechanisms (Figure 1.3). These models propose an interaction between the ventral and dorsal pathways so that the ventral stream can be responsible for solving the lack of invariance problem in the acoustic to phonetic mapping of incoming speech while the dorsal stream can represent demands from attention and the intentions of the listener for domain-general linguistic processes (Rauschecker & Scott, 2009). Importantly, they emphasize the role of top-down knowledge-based contextual cues in the feedback and feedforward interactions that occur between the two pathways in order to effectively interpret the message of the input received. Friederici (2012) even includes additional parallel ventral and dorsal pathways to accommodate the demands of syntactic complexity and cognitive control mechanisms. Similarly, Davis and Johnsrude (2007) add complexity to the model by asserting that multiple lexical interpretations of the speech signal are active at increasingly complex levels of representation to allow top-down projections from the inferior frontal, motor and somatosensory cortices to retune the perception of the acoustic signal at all cortical levels of analysis. With this addition to the network, knowledge from several sources is engaged to better interpret the incoming signal at multiple stages of the neural pathway.

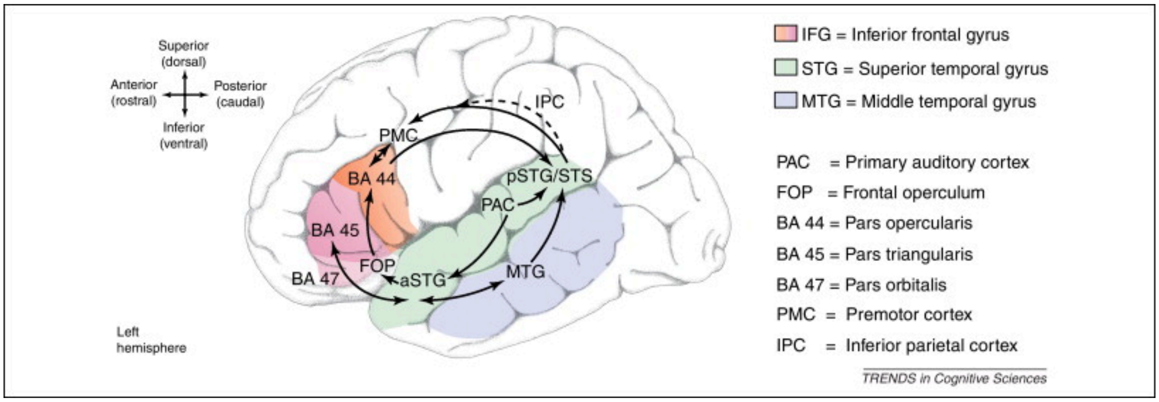

Figure 1.3. An example of the perisylvian language network proposed by several different neural accounts of speech perception. From Friederici (2012).

These models reflect an active top-down driven system. Motor theory describes speech perception as a process in which knowledge of speech production allows the speech network to create hypotheses about the incoming acoustic input. The neural models devised from motor theory reflect this process by illustrating a feedback connection between articulatory/motor movement areas and the structures that process the acoustic input. However, top-down driven active processes are not limited to a particular modality or knowledge base from which hypotheses about the input may be drawn. Both in claiming that these contextually-based processes are specific to speech perception and in limiting them to the perisylvian language network, motor theory and these neural models have failed to fully describe the flexible perceptual process that occurs when understanding sounds, which must include the influences of non-perceptual functions, such as attention, working memory and executive function. Additionally, these models assume that the acoustic signal is transmitted to the temporal cortex in a sequential, passive manner, ignoring increasing evidence that top-down projections can and do reach the subcortical levels of the peripheral nervous system.

As an alternative to motor theory based neural models, Skipper (2014) describes an account of speech perception that relies on a hypothesis driven top-down process outlined by the analysis by synthesis theory. Stevens and Halle (1967) proposed an analysis by synthesis theory based on a similar approach to that of motor theory. They hypothesized that acoustic signals are interpreted as articulatory configurations through a feedback loop of sorts in which stored articulatory models are compared with the incoming speech spectra. Multiple stored models are implemented until the model with the least number of errors when compared to the incoming signal is determined. The same conversion process is then engaged for each higher level of speech recognition, such as phonemes, words, etc. (Stevens, 1960). Based on this cognitive model, Skipper (2014) proposed the natural organization of language and the brain (NOLB) model to highlight the role of primary auditory cortex (A1) in speech perception. Unlike the neural models described above, Skipper argues that A1 does not simply engage in the basic transmission of the auditory signal, but rather takes part in the categorization of auditory objects by comparing the signal with hypotheses that it has received from several dynamic and simultaneously activated cortical networks. These hypotheses are derived from various cues that occur naturally within speech, such as mouth movements and gestures, as well as contextual cues that can be drawn from the environment and prior experience with the situation. Such cues are used to derive hypotheses about the incoming acoustic signal, which are transmitted to A1 to be compared with the actual auditory input. For situations in which many contextual cues are presented with the incoming speech, hypotheses are formed by cortical structures in the posterior ventral frontal (PVF) regions and transmitted to posterior superior temporal regions as potential auditory objects in an inverse model pathway. In instances when less contextual information is available, however, a forward model pathway transfers the incoming auditory signal from PST regions to PVF regions to be processed and understood using knowledge of speech production.

Of all of the neurobiological descriptions of speech perception, the NOLB most closely resembles a top-down hypothesis driven process in which a knowledge base is utilized to create multiple hypotheses about the perceptual representation of the incoming signal. It accounts for an interaction between multiple cortical networks so that both contextual and experiential factors may be integrated with the acoustic cues of the signal itself. Importantly, the NOLB asserts that hypotheses are derived for multiple levels of linguistic representation, so that feedback from several cognitive networks can have an impact on the signal as early as primary auditory cortex, an area that has largely been considered to be a passive transmission structure in the auditory pathway by other neural models of language processing. However, there are still aspects of the model that are problematic. Since the hypotheses of perceptual representations activated by the incoming signal are dependent upon prior knowledge of the situation as well as knowledge of speech production, the system can only recognize signals that it is able to synthesize itself, meaning it can only process input that it has experienced before. The NOLB cannot account for instances in which new stimulation is encountered and interpreted. Moreover, while the portrayal of primary auditory cortex as an initial site for a comparison between the acoustic representation and perceptual hypotheses begins to explain top-down influences on the early transmission of acoustic input, it still does not go far enough down the auditory pathway to fully explain the changes that can be seen along peripheral nerve pathways in response to changes in the internal states of the listener.

Each of these neural models of speech perception, despite proposing some variation of a top-down driven system that utilizes prior knowledge and contextual cues to derive hypotheses about the incoming signal, fail to fully address two important questions concerning this process: How far down the auditory pathway can these top-down projections have an effect on signal transmission and what kind of control system is employed by these projections? The extent of the top-down driven effects can be inferred from the assembly of the auditory pathway itself. As Figure 1.1 illustrates, efferent projections from the auditory cortex extend to most of the structures of the early sensory pathway, including the cochlear nucleus, suggesting that top-down projections from various cortical networks could potentially influence the encoding and transmission of acoustic signal at its earliest levels.

The type of control system that is incurred through these top-down projections is less clear. On the one hand, the emphasis on the motor system as a hypothesis generator in most neural models of speech perception suggests that the motor system may act as the main form of top-down control on the perceptual system by suppressing the processing of input in sensory brain regions so that motor hypotheses may drive the final interpretation of the input. Work by Skipper et al. (2007) examining the McGurk effect supports this form of motor-controlled perception by demonstrating that the final percept of a McGurk stimulus is driven by the motor representation of the input, a representation that is not based on the auditory or visual input but rather is a completely different percept that does not reflect either form of input. While the initial representation of the stimulus in the auditory and visual cortices reflects the sensory input of each system, the final representation reached by each neural network reflects the percept hypothesized by the motor system, indicating that the motor system may have driven the final percept reached by the visual and auditory cortices through a form of top-down, hypothesis driven control.

Evidence from speech production research also suggests that the motor system may act as a top-down control network for at least primary sensory regions. Houde and colleagues (2002) have demonstrated that the auditory cortex is suppressed for self-produced speech, but not external renditions of self-produced speech, indicating that the motor system may be actively suppressing the auditory cortex during vocal production. Similarly, suppression of the auditory cortex can be seen for self-triggered sounds, but not externally triggered sounds (Martikainen, 2005). This phenomenon is often explained as the consequence of a feed-forward control model between the motor system and sensory systems in which a corollary discharge (Sperry, 1950) or efference copy (von Holst, 1954) of the movement that is about to be produced is sent to a primary sensory area so that it may predict the consequences of that movement and then compare the prediction to the actual sensory outcome that is perceived. If there is a discrepancy between the predicted outcome and the actual outcome of the movement, an error signal is sent to the motor system through a feedback loop so that adjustments to motor production can be made (Blakemore, Wolpert & Frith, 2000). The consequence of this system is a suppression of certain neural populations in the auditory cortex so that the incoming signal from the movement that is transmitted by the auditory pathway does not override the corollary discharge signal received from the motor cortex. Extrapolating this mechanism to the realm of perception, the activation of the motor cortex to generate knowledge-based hypotheses about the incoming signal may incite a top-down control system in which these corollary discharge pathways are activated to suppress sensory systems so that the motor generated hypotheses may be given more weight in the perception process.

Limiting perception to solely sensory and motor networks, however, may be an incomplete portrayal of the top-down driven system that guides perception. Perception is a multi-dimensional activity that involves the representation of incoming signals, the analysis of input and the control of behavior in response to that input (Desimone & Duncan, 1995). Competition for attention resources exists at multiple levels of processing, from the transmission of relevant stimuli and suppression of irrelevant input in the environment to the activation of appropriate hypotheses about the incoming signal and inhibition of irrelevant hypotheses based on prior knowledge, expectations and pragmatics. This competition must be mediated in some way to ensure relevant stimuli are processed and the appropriate top-down knowledge is applied to create the correct percept. Desimone and Duncan (1995) proposed that an attention template, a description of the information required to perform a behavior based on the relevant task and situational cues, mediates the competition between these neural mechanisms in the visual system, however, their model could be applied to auditory perception as well. An attention template that arbitrates competing bottom-up auditory inputs as well as multiple knowledge or expectation-based hypotheses about the incoming signal, be they from motor, sensory or other cortical circuits, would best explain the top-down projections that are likely involved in auditory perception, as it would compare feedforward and feedback projections from all active systems to determine the percept that best fits the situation. Such a template would mediate the shifting of competitive weights between the demands of all active neural networks, making it independent from and yet inherently linked to sensory and motor networks.

Support for a top-down control system mediated by an attention template can actually be drawn from the same lines of research that advocate for a corollary discharge explanation. As described above, the McGurk effect has been characterized as a product of an efference copy of the prediction about the signal produced by the motor system constraining the processing of sensory input in the auditory and visual networks. The prediction made by the motor system appears to drive the final percept reached by the sensory systems so that they too interpret the input in the way that the motor system hypothesized (Skipper et al., 2007). However, Alsius et al. (2005) demonstrated that the effects of the motor system on audiovisual integration in the McGurk effect can be diminished under visual and auditory attentional load. In instances where auditory or visual attentional resources were divided between two tasks, the McGurk effect was less apparent, indicating that the mediation of competition between multiple systems and inputs plays an important role in the perception process. Likewise, production experiments that have been used to support a motor-controlled top-down process could be explained as a reallocation of attention to certain inputs over others that is mediated by the demands of the task at hand. Examinations of the motor suppression of the primate auditory cortex during self-initialized vocalizations reveal that the onset of this suppression occurs as early as 250ms before the onset of the vocalization and continues through the duration of the motor activity (Eliades & Wong, 2002). This early suppression of the auditory cortex indicates that what may appear to be motor inhibitory suppression of the auditory cortex may in fact be a preparatory reallocation of attention to the motor system in anticipation of the motor action that is about to occur.

Schneider, Nelson and Mooney (2014), while interpreting their results as further support for a corollary discharge explanation of top-down control, provide evidence that could support the idea of an overarching control system that modulates the allocation of attention as well. The authors examined the specific neural connections engaged in the suppression of the auditory system by measuring neural activity in the motor and auditory cortices of mice. Intracellular recordings of excitatory neurons in the primary auditory cortex during various types of movements, such as locomotion, head turns and vocalizations, revealed that excitatory auditory cortical neurons are suppressed for all forms of movement, not just movements that explicitly incur an auditory response. Importantly, the suppression of the auditory neurons begins as much as 200ms prior to the movement, continues throughout the motor action and ends up to 200ms after the movement has ceased. Further experiments demonstrated that the specific connections between the two cortices involve projections from M2 neurons in the secondary motor cortex onto PV+ interneurons, which then synapse on the excitatory auditory neurons in an inhibitory manner. Thus, when M2 neurons are activated by motor preparation, a sequence of events commences that results in a suppression of specific neural populations in the primary auditory cortex.

Schneider and colleagues' work suggests that the suppression of the auditory cortex during motion can occur in response to any form of movement. However, it is unclear whether this system can be fully explained by top-down control projections from the motor network. The authors admit that both M2 neurons and excitatory neurons in primary auditory cortex are components of much more complex networks, and the interaction between these two groups may not be as simple as their data suggest. The extended timing of the auditory suppression that they find compared to that of the motor movement suggests that this interaction between the motor cortex and the auditory cortex could be one small piece of a larger network in which the demands of the situation guide the amount of attention that is devoted to specific neural processes. If a motor movement inherently activates a series of feed-forward and feedback processes to ensure that the movement has been executed properly and has elicited the appropriate sensory response, sufficient amounts of attention must be assigned to both the motor system and the auditory system to achieve all of these processes. If the action and the sensorimotor consequences of that action are more important to the demands of the current situation than the acoustic consequences of the action, then attention must be distributed in such a way as to give more weighting to the motor event than to the auditory input. This type of interaction cannot be fully explained by a model that only includes direct, unidirectional projections between the motor cortex and the auditory cortex. It requires the consideration of situational and behavioral demands that could mediate the interaction of other cortical networks with these two systems, the way in which these two systems interact with each other, or the way in which the neural circuits, as a whole, processes and interact with the environment.

This dissertation addresses the questions concerning the top-down perceptual control system outlined above. To examine both the extent and the nature of these top-down projections, the auditory brainstem response was measured under various motor and attention conditions, as the response of the brainstem to auditory input appears to be influenced by top-down cortical processes in various ways. Electrophysiological recordings of the auditory brainstem response have demonstrated that the Frequency Following Response (FFR), a sustained response phase-locked to the fundamental frequency of a periodic stimulus and/or the envelope of the stimulus (Krishnan, 2007), reflects changes in higher-level cognitive processes such as attention and learning. For example, the frequency following response to synthetic English vowels contains prominent spectral peaks at the first formant harmonics of the signal and smaller peaks at the harmonics between formants (Krishnan, 2002). The enhanced peaks found in the FFR at the first formant suggest that some form of categorization based on prior experience and knowledge of speech input is already occurring at the level of the auditory brainstem. The representation of the input is modified to strengthen the important cues so that they are more prominent than the rest of the signal, indicating that the translation of signal to lexical representations that occurs in the cortico-cortical connections of the speech network may influence the way in which the auditory brainstem represents the signal as it is transferred to higher points along the auditory pathway. Similarly, Krishnan and colleagues (2005) compared the frequency following responses of Mandarin speakers and English speakers to four lexical tones used in Mandarin. They determined that Mandarin speakers had stronger pitch representations and smoother pitch tracking in their FFRs than English speakers. Based on these results, the researchers concluded that language experience may induce changes in the transfer of auditory input by the brainstem to enhance the representation of relevant linguistic features inherent in the signal. The interaction between experience and brainstem activity is not exclusive to language experience either. Musically trained individuals show earlier and larger FFRs and better phase-locking to the fundamental frequency in response to music stimuli as well as speech stimuli from their native language (Musacchia et al., 2007; Wong et al., 2007).

These studies demonstrate the long-term effects of perceptual experience and training on the responses of the auditory brainstem, a relatively low-level brainstem structure. Despite the fact that such effects are currently outside the domain of neurobiological theories of human perception, these data can reflect the role of a top-down control system in auditory speech perception, as they demonstrate descending experiential effects on the processing of speech and other acoustic stimuli. The top-down control of subcortical structures allows the system to be flexible and to adapt to the many internal and external changes that occur during perception of auditory and speech input by enhancing the most relevant spectral cues in the auditory input before it even reaches cortical perceptual networks. However, the evidence discussed thus far has only demonstrated top-down influences on brainstem structures after an extended period of training or experience, such as exposure to a tonal language or years of musical training.

Furthermore, it does not explain the nature of these top-down influences, whether they originate from a specific network, such as the motor system, or reflect the general distribution patterns of attention based on the environmental input and behavioral demands of the situation. In order to fully explain the nature of descending projections from cortical control systems, investigations of the auditory brainstem response must explore how these cortical networks may affect transmission of the acoustic signal by the brainstem in real time. They must also examine the type of top-down control that is being exerted over the system, be it a consequence of hypothesis derived activity in the motor system or a product of a system-wide process responsible for keeping track of inputs from the environment as well as the potential behaviors elicited by those inputs and distributing attention accordingly. Such data would further explain the extent to which these top-down descending projections influence processes in the human peripheral auditory pathway and provide additional information about the type of flexible neural model that best explains the complex process of perception.

The current research aims to understand whether cortically initiated phasic signals can modify ascending acoustic input in real time. To do so, the first two presented experiments make use of the prior observation that cortically initiated intentional motor behavior reduces sensory responses in the cortex (Schneider et al., 2014; Houde et al., 2002). The classic interpretation of this interaction is that corollary discharge projections from the motor system inhibit auditory system responses (Sperry, 1950; von Holst, 1954; Blakemore et al., 2000). The current research investigates whether phasic changes in cortical demands affect brainstem processing in the ascending auditory system. In other words, are the cortical suppression effects also manifest in earlier sensory mechanisms in the auditory brainstem? Experiments 1 and 2 test whether motor suppression of sensory processing in cortex also changes low-level neural representations of auditory signals that occur at the same time as the motor act by comparing the frequency following response to a simple and a complex auditory stimulus during production of intentional finger movements and a condition with no motor movement.

An alternative interpretation to the corollary discharge explanation is that a focal motor behavior demands attention and sensory activity is attenuated as a result of this demand (e.g. Newman et al., 2007). The demonstration that motor behavior changes auditory responses in the brainstem does not distinguish between corollary discharge and attention demand explanations. In a corollary discharge explanation, two conditions that have equivalent motor behavior (same number of movements with equal motor effort per movement) should produce similar reductions in auditory stimulus information in the FFR. By contrast, if the planning of the same number of movements differs and this difference in planning affects the FFR in spite of equal movements, this would be evidence that the suppression of the brainstem response might be a result of situational differences in attentional demand rather than corollary discharge. Experiments 1 and 2 address these two explanations by distinguishing two conditions with intentional motor behavior that contain the same number of induced finger movements with equal motor effort, but differ in the attentional demands of the task. Will two motor tasks that vary in attentional demand but have equal motor behavior produce different effects on the neural representation of auditory input in the brainstem?

Differential effects on the FFR between two conditions that have the same amount of motor behavior but differ in contextual attention demands may still be the result of corollary discharge projections from the motor system, as an interaction between the attentional demands and constant motor response could potentially arise due to changes in the corollary discharge reflecting load effects on the motor system. Experiment 3 varied the demands of a visual attention task while keeping the motor response constant. Will increasing the perceptual demands of a visual attention task modify the frequency following response to an acoustic target as well? If so, this would be additional evidence of top-down control over the descending auditory pathway based on a system that monitors and adapts to situational and behavioral demands.

Chapter 2

Experiment 1: Motor Responses Elicit Attentional Enhancement of the Brainstem Frequency Following Response to a Sine Tone

Human sensory systems, such as audition and vision, are often described as bottom up peripheral pathways that feed into cortical perceptual networks (Wolfe et al., 2012). Subcortical structures are portrayed as relay stations in which certain properties of the input signal, such as pitch or timing in the case of audition, are transmitted as a neural signal that can be subsequently recognized and processed by primary and secondary cortical sensory regions. Despite strong anatomical evidence for efferent projections in the peripheral auditory pathway (Garinis, Glattke, & Cone, 2011), most neural accounts of audition do not address how the cortical auditory network, or even complex cognitive networks, may influence the transmission of acoustic signal in the peripheral auditory pathway or the functional role in hearing of this possible influence.

Evidence from measurements of the frequency following response (FFR), the phase-locked component of the auditory brainstem response, suggests that complex cognitive processes can influence the way in which the brainstem represents pitch information. Tonal language speakers and highly trained musicians show better pitch tracking and larger spectral peaks in their FFRs to musical input as well as to speech input (Musacchia et al., 2007; Wong et al., 2007; Krishnan et al., 2005), suggesting that prolonged experience and training to attend to pitch information in the acoustic signal leads to an emphasis on the representation of the pitch component in all forms of meaningful acoustic input in the brainstem. Real time interactions between attention and the auditory brainstem response, however, are much less clear. Endogenous allocation of spatial attention to dichotic auditory input elicits disparate and inconsistent effects in the FFR (Galbraith et al., 1998; Galbraith & Arroyo, 1993; Hoormann et al., 2004). Similarly, the reported effect of general selective attention on the auditory brainstem differs across studies (Hoormann et al., 1994; Hairston et al., 2013; Galbraith et al., 2003), suggesting that the current methods used to examine the interaction between cortical networks and brainstem structures may be inadequate for properly isolating the frequency following response from other auditory processes or from engaging selective attention processes appropriately.

An examination of the interaction between the auditory brainstem and the motor system may be more suited for investigating the immediate influence of complex cortical structures on the peripheral auditory pathway. The auditory brainstem is linked to the motor system structurally, as the inferior colliculus contains projections to the cerebellum via the dorsolateral pontine nuclei, a connection that is important to the motor execution pathway (Hashikawa, 1983). At a cortical level, the link between the motor network and the auditory system has also been well established. Primate vocalization and speech production research has determined that the primary auditory cortex is suppressed during these types of motor behaviors (Houde, 2002; Eliades & Wong, 2002). Schneider and colleagues (2014) further demonstrate that, in mice, various forms of motor behavior lead to the suppression of a population of primary auditory neurons. It is important to note that these interactions are typically explained in terms of cortico-cortical interactions that do not involve subcortical structures (see Sperry, 1950; von Holst, 1954; Blakemore et al., 2000).

The relationship between the cortical motor system and the peripheral auditory pathway, however, has not been investigated systematically. Investigations of the auditory pathway across multiple species have demonstrated that self-initiated vocalizations are usually associated with attenuation of activity in subcortical structures. For example, decreased activation in the lateral lemniscus of the bat can be seen during vocalizations (Suga and Shimozawa, 1974) as well as in the human brainstem (Papanicolaou et al., 1986) during speech production. It is assumed that the attenuation of activity in these structures during vocalization acts to reduce the intensity of the utterance as it is encoded by the sensory system, through a form of corollary discharge. However, the connection between motor behavior and the transmission of acoustic input by the peripheral auditory pathway may not be as simple as the corollary discharge explanation suggests. Individual differences in the FFR to a syllable, /da/, predict differences in one's ability to tap in synchrony with a beat, with increased variability in the brainstem response correlating with an increase in tapping synchrony variability (Tierney & Kraus, 2013), suggesting that the relationship between the cortical motor network and the auditory system is much more complex and bidirectional than a corollary discharge explanation would allow. Given this evidence, it is possible that the suppression of cortical auditory processes may in fact be the product of a more widely distributed network that is responsible for monitoring the demands of the situation and allocating attentional resources throughout the system based on contextual and behavioral factors.

Based on the structural and functional link demonstrated between the cortical and subcortical motor system and the auditory system, this experiment investigates whether suppressive effects can be seen in the auditory brainstem in response to motor behavior as have been observed in the cortical auditory system. The FFR to a 440Hz sine tone was measured while participants engaged in two different finger-tapping conditions. These were compared to a Resting condition in which participants sat quietly while listening to the sine tone. To further determine whether any observed suppressive effects on brainstem activity are mediated by the demands of the task, the two finger-tapping conditions varied in difficulty. In one condition

(40bpm condition) participants were instructed to tap a photocell in response to every flash of a blinking light presented at 40 beats per minute. In the second condition (80bpm condition) participants tapped in response to every other flash of a light presented at 80 beats per minute, thereby equating motor behavior but varying cognitive demand. If the suppression of the auditory brainstem during motor behavior is the result of descending corollary discharge projections from the motor network, then there should be a general decrease in FFR spectral power at 440Hz in both tapping conditions compared to the Resting condition. However, if the auditory brainstem is suppressed in response to the increased task difficulty, which might increase demands on attention, then there should be different effects on FFR peak power for the two tapping conditions. Since the 80bpm condition requires more attention than the 40bpm condition, the representation of the pitch of the acoustic input in the auditory brainstem, a process that is not required to perform the tapping task, should be affected by the 80bpm condition compared to the 40bpm condition.

Materials and Methods

Participants

Fifty-four adults participated in this study; however, data was analyzed for only 47 of those participants. All remaining participants had normal bilateral audiometric thresholds (air conduction thresholds < 40dB HL for 500-4000Hz) and no history of learning or neurological disorders. All participants were compensated for their time with either experimental course credit or monetarily. Of the 7 participants who were excluded from data analysis, 3 did not show normal bilateral audiometric thresholds, as they failed to perceive all tone frequencies tested at 40dB HL in both ears, 3 were excluded due to equipment malfunctions during the study session and 1 participant showed no evidence of pitch representation in the FFR.

Primary Task

Three experimental conditions were administered; two tapping conditions and a Resting condition. In each of the tapping conditions, participants were asked to tap a finger against a photocell in response to a flashing light on a metronome. In the 40bpm condition, participants were presented with a visual metronome that flashed a light at 40 beats per minute and participants were instructed to tap the photocell each time they saw the light flash on. In the 80bpm condition, the metronome’s flashing light was set to 80 beats per minute and participants were instructed to tap the photocell every other time that they saw the light flash on. Thus, the rate at which participants tapped the photocell was the same (40 taps per minute) but the rate at which the light on the metronome flashed and the way in which the participants were instructed to respond to these blinking lights differed between the two conditions. In the Resting condition, participants did not engage in a tapping task, did not see a flashing light and were told to sit quietly while the tones were presented.

Background Tones

During all three experimental conditions a background auditory stimulus was presented to elicit the frequency following response within the auditory brainstem. A 440Hz pure tone (200ms in duration, 5ms linear amplitude ramps at on/offset) was presented at a rate of about 3 per second with a variable interstimulus interval of 100-300ms. Participants were told not to worry about the background tones and to focus on the tapping task in the two tapping conditions. Within the tapping conditions, the pure tone was presented in blocks of 1000 repetitions with a 2-minute break between blocks, at which time participants were not required to perform the tapping task. During the Resting condition, participants were told to relax and close their eyes if they wished, but they did not have to actively listen to the tones.

Procedure

This experiment was conducted over two sessions occurring on separate days. The two sessions were not required to occur on consecutive days nor at the same time each day. In one session, participants completed one tapping condition as well as the Resting condition. In another session, participants completed a tapping condition. At the end of each primary task condition, participants were asked to rate how difficult they found the task to be on a scale of 1 to 5, with 1 being not difficult at all and 5 being very difficult. Order of tapping condition as well as combination of tapping condition with Resting condition in a single session was completely counterbalanced. In the first session, participants completed a Musical Experience Questionnaire, to assess years of musical experience, as well as a basic demographics questionnaire to determine their age and native language(s).

Electrophysiology

Stimulus presentation

All brainstem recordings were conducted in an electrically shielded, soundproof booth. Brainstem EEG recordings were collected while background tones were presented binaurally via Etymotic Research ER-3A insert tubephones at 65-75dB at the ear. Stimuli were presented at alternating polarities to eliminate the presence of the cochlear microphonic in the recorded brainstem response. Each polarity of the background tone was presented a total of 1500 times (3000 total renditions of the tone) in each condition of the experiment. For the tapping conditions, the presentation of the stimulus was broken up into three smaller blocks of 1000 presentations each to give participants a break from the tapping task.

Recording and data processing procedure

Brainstem responses were collected at a sampling rate of 25kHz using scalp electrodes and a BrainVision actiChamp amplifier system (BrainProducts, Munich, Germany). Four Ag-AgCl electrodes were placed on the scalp in a vertical montage (Hood, 1998), with the active electrode placed at central vertex (Cz), linked reference electrodes placed on each earlobe and a ground electrode on the forehead. Contact impedances were kept under 5k Ohms. Recordings were made in BrainVision PyCorder software (BrainProducts) in continuous mode with an online filter of 0.1-3000 Hz. Data from the photocell were also recorded by the actiChamp system at the same sampling rate to monitor the tapping motions of the participants.

All preprocessing of the neurophysiological data was conducted using BrainVision Analyzer 2 software (BrainProducts). For analyses conducted on the brainstem response, recordings were bandpass filtered off-line from 100 to 2000 Hz (12 dB/octave rolloff) using a Butterworth filter. A 60Hz notch filter was also applied to the EEG recordings to eliminate line noise. Each tone was epoched with a window of -50 to 250ms. Artifacts were rejected at a +/-35µV voltage threshold. On average, less than 1% of segments were discarded due to artifact. The remaining pool of artifact-free epochs were baseline corrected using an interval of -50 to 0ms, averaged within each condition and visually inspected for the presence of a frequency following response and a high signal to noise ratio. The averaged waveforms were exported for further analysis, converted to wav files in MATLAB (The MathWorks, Natick, MA) and then the phase-locked component of the brainstem response was examined in Praat (Boersma, 2001). Spectral slices were taken of each wav file and then the peaks of the spectrum tier were calculated to determine the amount of spectral power (dB) present in the FFR at 440Hz.

Averaged waveforms for each block within each tapping condition were also extracted to monitor quality control of the FFR throughout the recordings. After using the same filter as described above, recordings for the two tapping conditions were segmented into three separate blocks, one for each block of recording between breaks, and all further preprocessing steps were equivalent to those described above. Repeated measures ANOVAs for each tapping condition revealed that there is no difference in the amount of spectral power at 440Hz in the FFR across blocks in the 40bpm condition ($F(2,46) = 0.996$, $p = 0.32$) or in the 80bpm condition ($F(2,46) = 6.4$ e-05, $p = 0.99$).

Tapping data from the photocell was also processed using the BrainVision Analyzer 2 software. Photocell recordings were bandpass filtered from 1 to 2000 Hz (12 dB/octave rolloff) using a Butterworth filter. The onset, peak and offset of each tap was marked based on change in signal amplitude above and below a baseline threshold. Tone presentations were then epoched based on whether one of these markers for tap onset, peak or offset occurred within a window of -50 to 250ms of the onset of the tone. After artifact rejection, baseline correction and averaging, using the same procedures as described above, similar spectral analyses were performed in Praat to compare the amount of power at 440Hz in the brainstem response for instances in which a tap and tone occurred concurrently with instances in which a tone was played in the absence of a tap.

EEG recordings were processed and analyzed to measure cortical EEG responses as well. In particular, frequency band analyses were performed to determine the strength of theta, alpha, beta and gamma oscillations within each condition. For these cortical EEG analyses, recordings were bandpass filtered off-line from 0.1 to 120Hz (12 dB/octave rolloff) using a Butterworth filter, along with a 60Hz notch filter. After segmenting out the break period between blocks within each condition, recordings were epoched into 1s intervals, an FFT with a Hanning window of 10% length was performed on each 1s epoch to determine frequency power and then those FFTs were averaged for all epochs. Maximum peak spectral power information for each frequency band was extracted from the averaged FFT and compared across conditions. Similarly, cortical oscillation data was extracted from the EEG signal to compare instances in which a tap and tone occurred together with instances in which a tone was played without the presence of a tap.

Results

As a first analysis, the Frequency Following Response (FFR) was compared across the three experimental conditions. Mean spectral power at 440Hz was 50.40 dB (SD = 5.44) for the Resting condition, 50.91 dB (SD = 4.82) for the 40bpm condition and 51.16 dB (SD = 4.56) for the 80bpm condition. A one-way repeated measures ANOVA revealed a non-significant main effect of condition ($F(2,92) = 0.014$, $p < 0.99$). When examining individual participants, however, the differences between the three test conditions varied greatly, with some participants showing suppression in response to the tapping conditions, while others showed enhancement. For example, when comparing the Resting condition to the 40bpm condition, 21 participants showed suppression of the representation of pitch information in the FFR, while 26 showed enhancement. Similarly, a comparison of FFRs in the Resting condition and the 80bpm condition showed that 20 participants showed suppressive effects and 27 showed enhancement effects. This relatively even distribution of enhancement and suppression across individuals was even evident when comparing the two tapping conditions, with 25 participants showing suppression and 22 showing enhancement. However, when looking at the differences between test conditions within the individual, only 15 participants showed consistent suppressive or enhancement effects across all three difference scores.

Years of musical experience was regressed with the differences in spectral power between experimental conditions (Resting – 40bpm, Resting – 80bpm and 40bpm – 80bpm). None of the three regressions yielded significant results (Resting – 40bpm: $R^2 = -.002$, $t(45) = 0.14$, $p > 0.9$; Resting – 80bpm: $R^2 = -0.01$, $t(45) = 0.62$, $p > 0.5$; 40bpm – 80bpm: $R^2 = -0.02$, $t(45) = 0.51$, $p > 0.6$), indicating that years of musical experience did not predict changes in the FFR between conditions.

While the effect of the motor task on the brainstem response is unclear both when comparing across the three experimental conditions and when looking across and within the individual, an effect of the tapping behavior may still be present in the two tapping conditions. Within the 40bpm and 80bpm conditions there are in fact two different types of events – those in which a tap occurred while a tone was presented (Taps) and those in which a tone occurred alone (No Taps). The underlying psychological processes and behaviors that are involved in each of these types of events are different, with No Taps events lacking the addition of a motor movement while processing the tone. Thus, rather than treating all auditory events within each tapping condition as the same, it is reasonable to examine whether it is not the introduction of a motor task overall, but the addition of a motor task to a specific auditory event that causes an interaction between the motor system and the peripheral auditory pathway.

As outlined in the data processing section, data from the 40bpm and 80bpm conditions were segmented into Taps and No Taps events. Taps segments consisted of instances in which a tap occurred within 50ms before or 250ms after tone onset. All other segments were considered No Taps events. As 3000 tones were presented in each condition, but participants only tapped the photocell at a rate of 40 taps per minute (once every 1.5s) in each tapping condition, the numbers of segments in the Taps and No Taps groups were not equal. On average, the 40bpm condition contained 869 Taps segments (SD = 115) and 2098 No Taps segments (SD = 127). Similarly, the 80bpm condition included 837 Taps segments (SD = 100) and 2143 No Taps segments (SD = 96). A 2x2 repeated measures ANOVA of spectral power at 440Hz in the FFR demonstrated a significant main effect of tapping ($F(1,46) = 10.99$, $p < 0.002$; Figure 2.1), with Taps events (Mean = 51.1 dB, SD = 4.47) showing an enhancement of peak power compared to No Taps events (Mean = 50.86 dB, SD = 4.65). The main effect of condition as well as the interaction of condition by tapping, however, was not significant ($p > 0.05$).

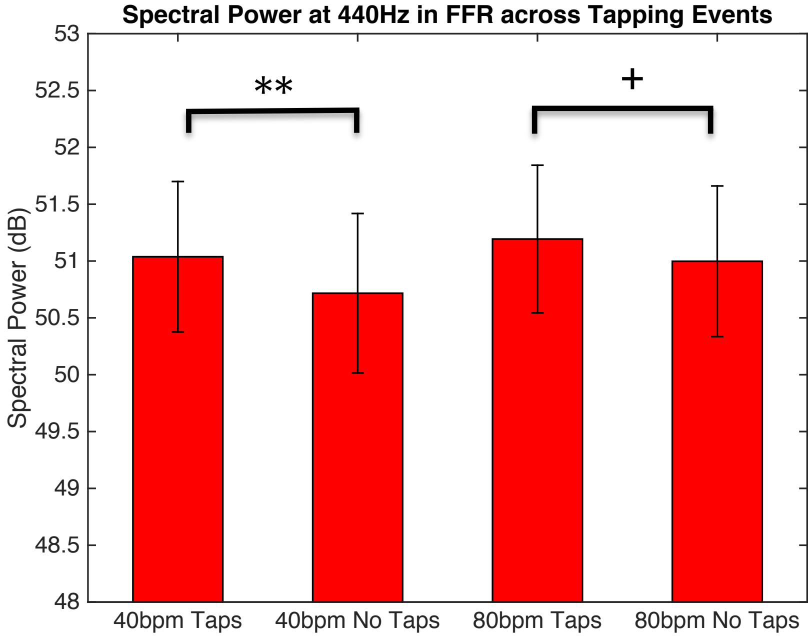

Figure 2.1. . Comparison of the spectral power at 440Hz in the FFR for Taps and No Taps events in the two tapping conditions (40bpm and 80bpm). Asterisk denotes statistical significance ($t(46) = -2.74$, ** $p < 0.01$) Difference between Taps and No Taps events in the 80bpm condition was marginally significant ($t(46) = -1.88$, + $p < 0.07$).

Since No Taps events included a tone presentation without the presence of a motor movement, one could consider them to be comparable to auditory events in the Resting condition, as all tones within this condition were presented without the occurrence of a tap. A two-tailed paired t-test comparing the Taps segments to the No Taps segments collapsed with the

Resting condition segments revealed that the spectral power at 440Hz for the Taps events was significantly larger than that for the No Taps and Resting events ($t(46) = -2.06$, $p < 0.05$; Figure 2.2).

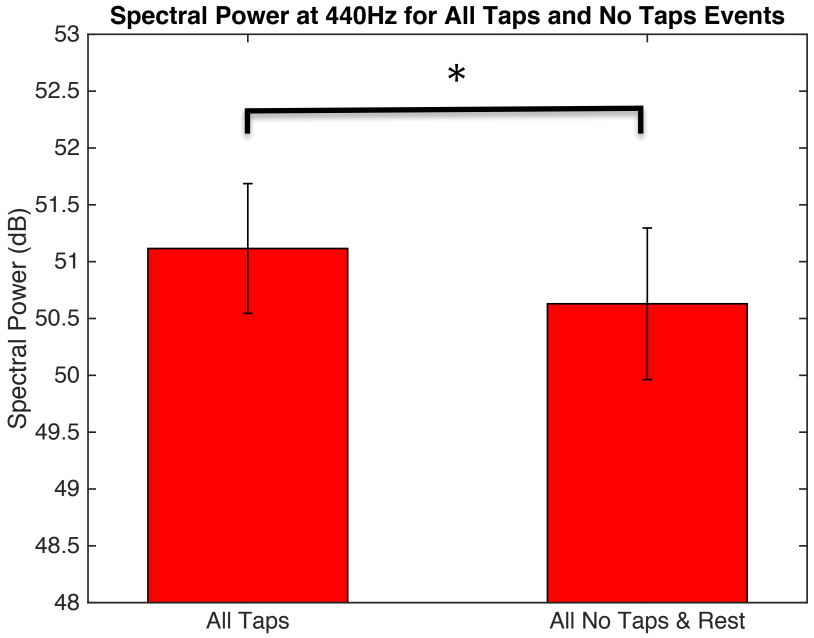

Figure 2.2. Comparison of the FFR for all Taps events with all No Taps and Resting condition events. The y-axis plots spectral power at 440Hz in the FFR in dB. Asterisks denote statistical significance (* $p < 0.05$).

To begin to understand the interactions that occur between cortical networks and the auditory brainstem, cortical oscillation measurements were extracted from the EEG recordings and compared across the three conditions of the experiment as well. Maximum peak spectral power was calculated for theta, alpha, beta and gamma bands in each experimental condition. In all four frequency bands, a one-way repeated measures ANOVA revealed a significant main effect of condition, with peak spectral power significantly larger in the Resting condition than in each of the two tapping conditions (Table 2.1).

Frequency Band	Resting Condition (μV^2)	40bpm Condition (μV^2)	80bpm Condition (μV^2)
Theta (4-8Hz)	Mean = 2.98** SD = 2.07	Mean = 1.98 SD = 1.08	Mean = 1.86 SD = 1.01
Alpha (9-12Hz)	Mean = 3.41** SD = 3.65	Mean = 1.96 SD = 2.41	Mean = 1.88 SD = 1.89
Beta (13-20Hz)	Mean = 0.59** SD = 0.49	Mean = 0.35 SD = 0.21	Mean = 0.35 SD = 0.22
Gamma (25-50Hz)	Mean = .15** SD = 0.08	Mean = 0.12 SD = 0.07	Mean = 0.12 SD = 0.05

Table 2.1. Maximum peak spectral power (μV^2) for theta, alpha, beta and gamma oscillation bands for each experimental condition. One-way repeated measures ANOVAs for each frequency band revealed a significant main effect of condition (theta: $F(2,92) = 11.46$, $p < 0.000$; alpha: $F(2,92) = 15.96$, $p < 0.000$; beta: $F(2,92) = 11.46$, $p < 0.000$; gamma: $F(2,92) = 10.24$, $p < 0.000$). Asterisk denotes statistical significance in paired t-tests with other experimental conditions (** $p < 0.01$).

Given the significant differences demonstrated in the FFRs between the Taps events and the No Taps events, theta, alpha, beta and gamma frequency band oscillations were also compared between these two event types. 2x2 repeated measures ANOVAs comparing peak spectral power for the two tapping conditions and the two tapping event types revealed a significant interaction effect of Condition x Event Type for both the theta ($F(1,46) = 9.09$, $p < 0.005$) and alpha ($F(1,46) = 5.288$, $p < 0.03$) frequency bands but no main effects of condition or event type ($p > 0.05$). This interaction effect is likely due to the fact that in the 40bpm condition, Taps events are significantly larger in peak spectral power than No Taps events (theta: $t(46) = -2.40$, $p < 0.03$; alpha: $t(46) = -2.41$, $p < 0.03$), but in the 80bpm condition maximum spectral power in the theta and alpha bands do not differ ($p > 0.05$; Figures 2.3 and 2.4).

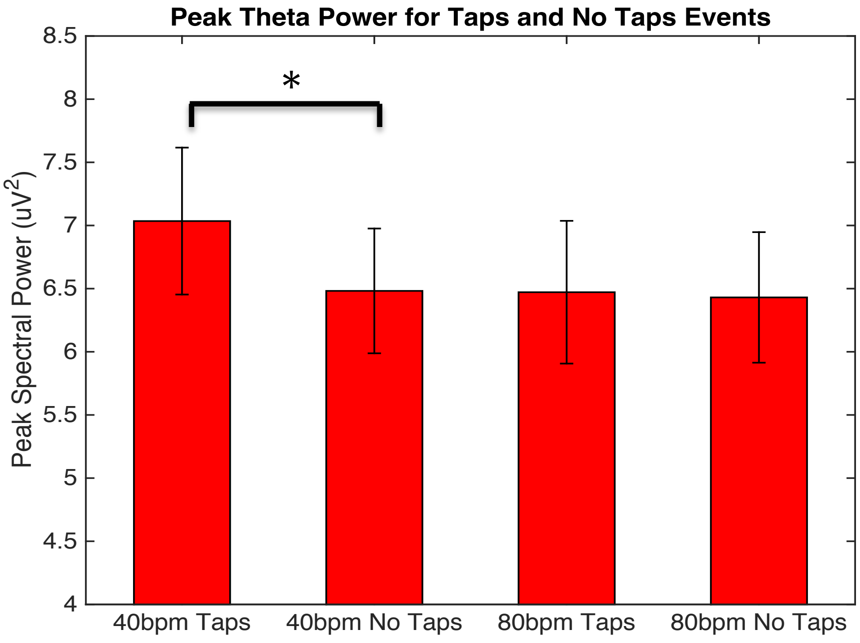

Figure 2.3. Comparison of peak spectral power in the theta frequency band for Taps and No Taps events in the two tapping conditions. The y-axis plots peak spectral power in (μV^2) at 6.10Hz. Asterisks denote statistical significance (* $p < 0.05$).

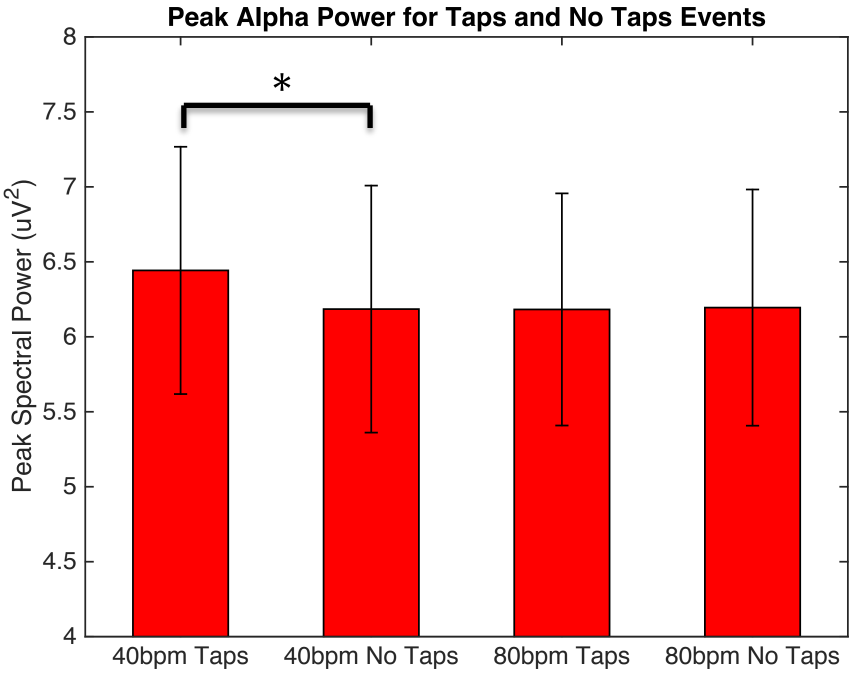

Figure 2.4. Comparison of peak spectral power in the alpha frequency band for Taps and No Taps events in the two tapping conditions. The y-axis plots peak spectral power in (μV^2) at 9.16Hz. Asterisks denote statistical significance (* $p < 0.05$).

In contrast with the results of the theta and alpha band analyses, 2x2 repeated measures ANOVAs comparing the peak spectral power in the beta and gamma oscillations bands for the 40bpm and 80bpm conditions and the Taps and No Taps events resulted in a significant main effect of event type for both beta ($F(1,46) = 15.48, p < 0.001$) and gamma ($F(1,46) = 31.28, p < 0.000$), but no significant main effect of tapping condition or interaction between tapping condition and event type. Subsequent paired t-tests between Taps and No Taps events demonstrated that No Taps events had significantly more peak spectral power in both the beta and gamma frequency bands for the 40bpm tapping condition (beta: $t(46) = 2.00, p = 0.05$; gamma: $t(46) = 4.78, p < 0.000$) as well as for the 80bpm condition (beta: $t(46) = 4.46, p < 0.000$; gamma: $t(46) = 4.99, p < .000$; Figures 2.5 and 2.6).

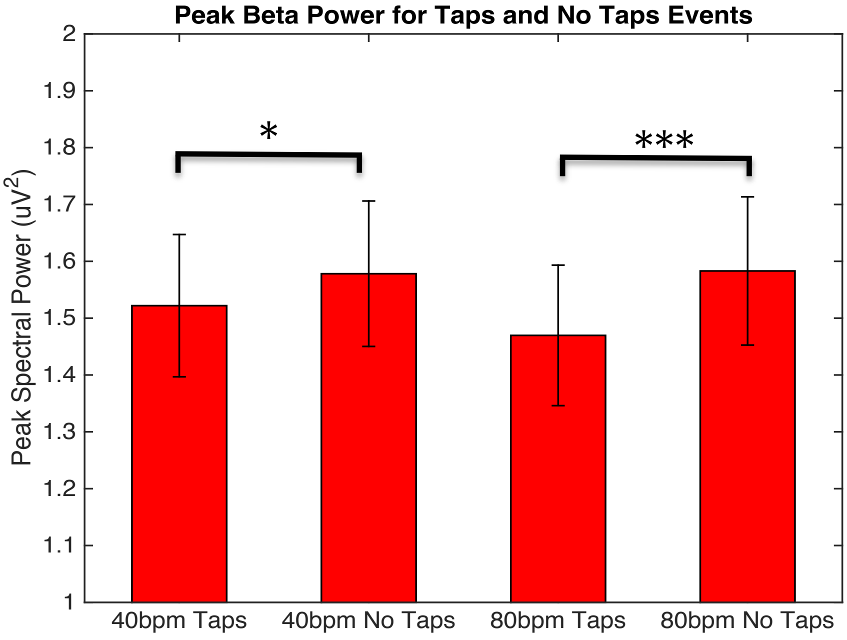

Figure 2.5. Comparison of peak spectral power in the beta frequency band for Taps and No Taps events in the two tapping conditions. The y-axis plots peak spectral power in (μV^2) at 15.26Hz. Asterisks denote statistical significance (* $p = 0.05$; *** $p < .000$).

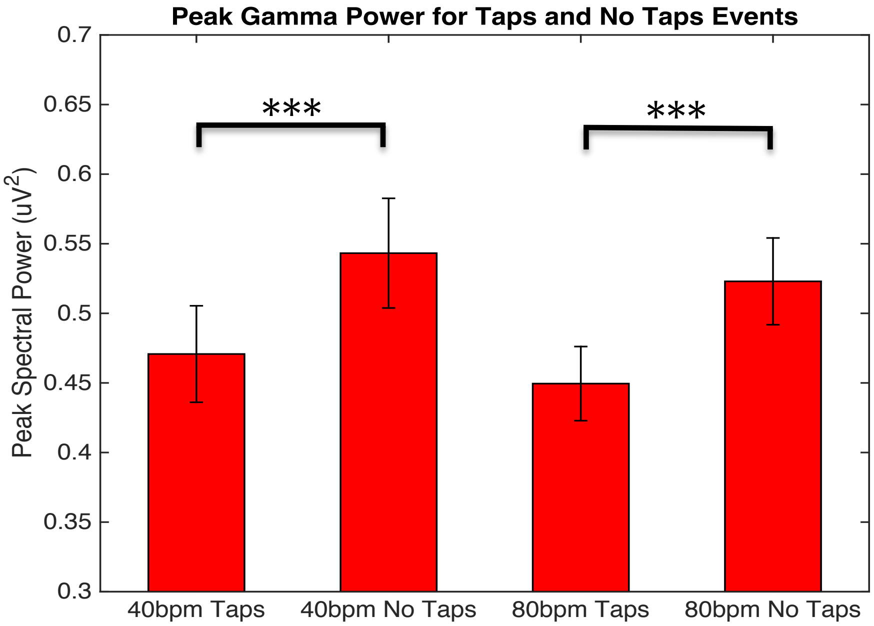

Figure 2.6. Comparison of peak spectral power in the gamma frequency band for Taps and No Taps events in the two tapping conditions. The y-axis plots peak spectral power in (μV^2) at 27.47Hz. Asterisks denote statistical significance (*** $p < .000$).

Discussion

The first aim of this experiment was to determine whether the motor suppression effects seen in the primary auditory cortex during motor behavior could be demonstrated at the level of the auditory brainstem. To that end, the results of this experiment do not support the hypothesis that performing a tapping task while hearing a sine tone will reduce the specific frequency following response to the tone. Overall, the amount of spectral energy at the frequency of the tone in the FFR did not differ between the three experimental conditions, even though individual differences between the three conditions demonstrated both suppression and enhancement of the FFR. For each comparison of experimental conditions (Resting compared to 40bpm, Resting compared to 80bpm, 40bpm compared to 80bpm) roughly equal numbers of participants showed enhancement and suppression of the representation of the tone in the FFR. However, these differences between conditions were not consistent within the individual.

A closer examination of the specific events within the two tapping conditions, however, begins to clarify how the introduction of a tapping task during the presentation of auditory stimuli affects the auditory brainstem response. When the events within the 40bpm and 80bpm conditions were separated into those in which a tone was presented while the participant tapped the photocell and events in which the tone was presented on its own, an enhancement effect of motor behavior on the FFR emerges. Representation of tone frequency was significantly larger for the Taps events than for the No Taps events, indicating that, rather than suppressing the response of the brainstem to the auditory stimulus, the motor behavior actually enhanced the frequency following response.

To further examine the interaction between cortical networks and the peripheral auditory pathway, EEG recordings were analyzed in terms of cortical frequency band oscillations as well. A comparison of theta, alpha, beta and gamma frequency bands across the three experimental conditions revealed a significant drop in peak spectral power between the Resting condition and the two tapping conditions for all four frequency bands. Further investigation into cortical EEG activity between the Taps events and the No Taps events, however, revealed important differences between these oscillations. While peak spectral power in the theta and alpha bands increased for the Taps events compared to the No Taps events (at least in the 40bpm condition), beta and gamma oscillation power was significantly larger for the No Taps events than for the Taps events, suggesting that various cortical networks may be engaged in different psychological processes during these two types of events.

The FFR results of this experiment are not only contrary to the predictions that motor behavior will suppress auditory brainstem responses and that attentional demands will further attenuate the brainstem response to auditory input, but also support a response-based attentional enhancement of the phase-locked component of the auditory brainstem response rather than a stimulus or condition-based effect. Research has demonstrated enhancement of the FFR to the same note patterns when they are presented in a pseudo-random order compared to a patterned order (Skoe et al., 2013) as well as to a synthetic syllable when it is presented repetitively compared to when it is presented with other speech sounds that vary in formant structure, duration, voice onset time or fundamental frequency (Chandrasekaran et al., 2009). This research displays contextual enhancement of the FFR between blocks, but not differences in the FFR on an event-by-event basis. Additionally, Skoe and Kraus (2010) showed that the repetition of a note within a melody that is repeatedly presented for 90 minutes can enhance the local representation of pitch information after prolonged exposure to a stimulus pattern, but listeners were not required to engage in a task while the melody was presented. The current experiment however, provides evidence for a response-based enhancement of the representation of acoustic input in the auditory brainstem in real time. The present results suggest that, rather than suppressing the neural representation of the auditory signal, the engagement of the motor system to perform the tapping motion may mobilize a set of processes that appear to direct and boost attention to the auditory signal during that specific event. This rejection of the corollary discharge process in the motor system is supported by the FFR findings of this experiment as well as several investigations of the effect of rhythmic stimulus presentation on auditory and visual perception.

Large and Jones (1999) proposed a dynamic attending theory of auditory perception in which they asserted that attention is not spatially selective, but temporally selective, meaning that the temporal structure of external events, not the location in which those events occurred, guides selective attention. They argue that internal oscillations within the individual entrain to the temporal structure of environmental events and generate expectancies for future events so as to enhance attention, and ultimately perception, within those specific time windows. Evidence to support this theory has been generated in both the visual and auditory realm. For example, Jones et al. (2002), demonstrated that pitch perception of a tone is more accurate when the timing of the presentation of the tone matched the rhythm of a series of distractor tones. Similarly, Mathewson and colleagues (2010) found momentary release of visual masking in a rapid serial visual presentation task when the masked target was presented at the same inter-stimulus-interval as the other visual stimuli in the task.

The results of this experiment further support the dynamic attending theory by demonstrating that the neural representation of pitch information is enhanced when an auditory stimulus is presented at the same time interval as an entrained visual stimulus. Throughout the two tapping conditions of this experiment, participants watched a metronome as it flashed at a constant rate of either 40 beats per minute or 80 beats per minute. In each condition, the tapping motion was elicited by a flash of the metronome light, making the two relatively in sync temporally. This flash of the light at consistent time intervals likely entrained internal oscillators to the timing of the light flash and tuned selective attention processes to the timing of each flash and the subsequent tapping motion, leading to an enhancement of the FFR to the tone, but only when it co-occurred with a tapping motion. Since the results of this experiment demonstrate an increase in neural pitch representation during a motor movement, the corollary discharge explanation for the link between the motor network and the auditory system cannot explain these findings. Given the rhythmic presentation of the visual stimulus that also elicited a rhythmic motor movement by the individual, theories supporting the temporal tuning of attention to specific events are more suited to explain these data.

The cortical EEG results from this experiment also support the idea that the peripheral auditory system is influenced by a temporally driven selective attention network. First, the pattern of results found in the theta and alpha band oscillations not only suggests that an attention network may be involved in the selective enhancement of the FFR, but also refutes the idea that descending corollary discharge projections from the motor network may play a role in this process. Examinations of the cortical motor system have demonstrated that motor behavior is characterized by suppression of alpha band frequencies in the motor cortex, an effect commonly referred to as mu suppression (Chatrian et al., 1959). While the comparison of alpha band power between the three experimental conditions exhibits this effect, with alpha power significantly decreased in the two tapping conditions compared to the Resting condition, the differences in alpha power for the two types of tapping events does not follow the same pattern. Alpha spectral power was actually increased for Taps events compared to No Taps events in the 40bpm condition. Coupled with the same pattern of results in the theta band oscillations, this evidence suggests that cortical networks engaged in selective attention and working memory processes influenced the auditory brainstem response during these specific events (von Stein & Sarnthein, 2000; Sauseng et al., 2005), not efferent corollary discharge projections from the motor system.

The pattern of results for the beta and gamma band oscillations further supports the role of selective attention in the enhancement of the FFR during Taps events. Spectral power in the beta and gamma bands was significantly larger for the No Taps events compared to the Taps events in both tapping conditions, contrasting with the findings from the theta and alpha band comparisons as well as the pattern of results in the FFR analysis. Such findings suggest that these two cortical oscillations were involved in the execution of the motor movement and were not involved in the enhancement of the perception and neural representation of the auditory stimulus.

Several experiments examining the activity of cortical beta and gamma oscillations in humans and non-human primates have demonstrated that beta and gamma oscillations are more synchronous in the time interval leading up to a motor movement and are suppressed during actual motor behavior (Sanes & Donaghue, 1993; Swann et al., 2009; Engel & Fries, 2010). Since beta and gamma power were largest when a tone was presented without a tapping behavior and significantly decreased when a motor behavior occurred, it is reasonable to conclude that this cortical activity may be linked only to the motor task required in the two tapping tasks and is not directly related to event-related enhancement effects seen in the auditory brainstem recordings.

The brainstem and cortical EEG findings of this experiment provide preliminary evidence that neural processes responsible for the focusing of attention at specific moments in time based on contextual factors influence the neural representation of the acoustic input by the peripheral auditory pathway. While this experiment cannot address whether these neural processes originate from cortical activity, the similarity in the patterns of results found for the brainstem response and cortical theta and alpha oscillations suggest that the changes seen in the FFR are related to changes in cortical neural activity in some way. However, several caveats of the paradigm and the results require further investigation into this phenomenon.

The overall difference in cortical oscillation synchrony between the three experimental conditions is slightly worrisome. While the decrease in alpha synchrony between the Resting condition and the two tapping conditions is to be expected, as the tapping conditions included a motor movement, the decrease in synchrony for the theta, beta and gamma bands was unexpected given the design of the experiment. The Resting condition, however, was a very boring task in which participants were instructed to simply sit quietly while the tones were presented for nearly twenty minutes. They were told that they were not required to pay attention to the tones and that they could close their eyes if they wished. Under these circumstances, most participants fell asleep during this condition of the experiment, which likely induced an "idling rhythm" within the lower cortical frequencies (1-20Hz) (Pfurtscheller et al., 1996; Steriade et al., 1993). To eliminate the effects of sleep on cortical oscillation synchrony while still obtaining a baseline measurement of pitch representation in the auditory brainstem, participants should be given a passive task to perform while the auditory stimuli are presented.

Additionally, the ecological validity of the auditory stimulus presented to the listeners should limit the scope of the conclusions drawn from this experiment. Sine tones are rarely, if ever, encountered in the world and they are the simplest form of sound that exists, making it difficulty to apply the results of this experiment to other forms of auditory input. Sine tones are also hardly ever produced by an action in the world, limiting the ability to compare this experiment to investigations of the suppression of the auditory system during speech production. Before further conclusions about the influence of the motor network, or any other cortical networks, on the peripheral auditory pathway can be made, the limitations of this experiment must be addressed.

Chapter 3

Experiment 2: Motor Responses Elicit Attentional Enhancement of the Brainstem Frequency Following Response to a /da/ Syllable

Experiment 1 offers two important pieces of evidence regarding the influence of efferent cortical projections on the auditory brainstem response. First, it demonstrates that the suppressive effects of corollary discharge projections from the motor cortex to the primary auditory cortex do not extend to the auditory brainstem. Moreover, the experiment provides evidence to suggest that a motor response to a periodically presented visual cue leads to enhancement of the frequency following response to a sine tone when the motor behavior and the auditory stimulus co-occur in time. Such results indicate that top-down cortical projections to the auditory brainstem may originate from a widespread attention system, as opposed to the motor system. However, in order to fully support this conclusion, the limitations regarding the auditory stimulus and the Resting condition of Experiment 1 must be addressed.

The primary aim of the first experiment was to determine whether motor suppression effects seen in the auditory cortex extend to the peripheral auditory pathway, particularly the auditory brainstem. The results of the experiment would suggest that this is not the case, as the simultaneous occurrence of a tap and a sine tone resulted in an increase in the phase-locked component of the brainstem response to the pitch of the tone rather than suppression. The use of a sine tone to elicit the brainstem response, however, may have resulted in a unique, and potentially ecologically invalid, situation for the listener. Most research examining the interaction between the motor system and the auditory system in humans and non-human primates presents auditory stimuli that either have been or could be produced by the listener and find that the auditory system is suppressed during motor activity (Houde et al., 2002;

Martikainen, 2005; Eliades & Wong, 2002). A 440Hz pure tone is hardly a sound that is commonly heard, let alone produced, by most human listeners. Musicians may be a bit more familiar with pure tones, but regression analyses did not produce a significant relationship between musical experience and the neural representation of pitch information in Experiment 1, suggesting that a sine tone may not even successfully elicit the motor suppressive effects seen in the auditory system at a cortical level, let alone at the level of the brainstem.

Additionally, previous research that has examined the effects of spatial selective attention on the auditory brainstem response argued that the relationship between attention and the FFR is largely mediated by the complexity of the auditory stimuli presented. Galbraith and Arroyo (1993), for example, found mixed effects of selective attention on the FFR to dichotically presented pure tones, with attention producing both enhancement and suppression in the FFR within the same recording session. Galbraith and Doan (1995), however, found stronger attention effects on the FFR in response to complex, but meaningless, stimuli containing a missing fundamental, and Galbraith and colleagues (1998) saw the most robust influence of spatial selective attention on the brainstem response when dichotic vowels were utilized in the task. This discrepancy in attentional effects on the auditory brainstem response based on stimulus complexity could explain why conditional differences were not seen between either the Resting condition and the two tapping conditions overall or between the 40bpm and the 80bpm conditions in Experiment 1. In other words, a 440Hz sine tone may not be complex enough to truly engage the cortical networks involved in either selective attention processes or suppressive corollary discharge activity.

To ameliorate these potential problems with engaging motor and attention networks, Experiment 2 attempted to replicate the findings of Experiment 1 by measuring the phase-locked component of the brainstem response to the speech syllable /da/. The use of a speech syllable instead of a pure tone should not only keep the amount of experience each participant has with hearing this type of stimulus relatively constant, as only English speakers participated in the experiment, but it should also fully engage the cortical motor system. Speech syllables are produced by the mouth and, according to many neural models of speech perception (e.g. Liberman et al., 1967; Rauschecker and Scott, 2009; Friederici, 2012), should activate, to some degree, the same motor areas involved in producing the syllables when they are perceived. The use of a complex speech stimulus should also eliminate any potential issues with the activation of selective attention networks when using simple auditory stimuli to elicit the brainstem response.

Experiment 2 is a replication of Experiment 1 with the auditory stimulus switched from a 440Hz sine tone to a synthesized /da/ syllable. Participants were once again asked to tap a photocell in response to the flashing of a metronome light, with participants tapping the photocell in response to every flash of the light in the 40bpm condition and in response to every other flash of the light in the 80bpm condition. While sleeping is often encouraged during most ABR experiments because the brainstem response remains robust during sleep (Galbraith et al., 2000), the stark contrast in arousal levels between the Resting condition and the two tapping conditions in Experiment 1 may have acted as a confounding variable in the comparison of the FFR between the three experimental conditions. Experiment 2 addresses this issue by having participants watch a silent television program or movie with subtitles during the Resting condition.

If the motor system suppresses the auditory system during motor acts, but only when the auditory system processes sounds that could be produced by the listener, then there should be a significant decrease in the phase-locked component of the brainstem response to the speech syllable between the Resting condition and the two tapping conditions, but no difference between the 40bpm and 80bpm conditions. However, if the efferent projections in the corticofugal system stem from attention processes, there should be a significant difference in the FFR between the Resting condition and the 40bpm condition, and a further difference in the FFR between the 40bpm and 80bpm conditions. These effects may be seen across all auditory events within each condition or when the events within the tapping conditions are further categorized into Taps and No Taps events. If the differences between these conditions are negative, meaning there is a decrease in spectral power at the pitch and first formant of the speech syllable in the FFR when the motor task is combined with the auditory task, this would indicate that cross-modal selective attention processes might be suppressing the representation of task-irrelevant acoustic information in the auditory brainstem. An enhancement of the FFR when the tapping behavior and the auditory stimulus co-occur, either in a certain experimental condition or in time, would suggest that an attention system involved in the temporal tuning of the perceptual system to the environment may be driving this effect.

Materials and Methods

Participants

Recordings from twenty-four adults were included in this experiment (2 participants were eliminated from data analysis due to equipment malfunction). All included participants had normal bilateral audiometric thresholds (air conduction threshold <40dB HL for 500-4000Hz), and no history of learning or neurological disorders. All participants were compensated for their time with either experimental course credit or monetarily.

Primary Task

The primary tapping task for this experiment was identical to that of Experiment 1. Two tapping tasks were administered in which participants tapped a photocell in response to every flash of a metronome light that flashed at 40 beats per minute (40bpm condition) or in response to every other flash of a metronome light that flashed at 80 beats per minute (80bpm condition). Once again, the rate at which the participants tapped the photocell remained constant across tapping conditions while the rate of the flashing metronome light varied. During the Resting condition, participants sat quietly and watched a television show or movie of their choosing with subtitles while the auditory stimulus was presented.

Background Auditory Stimuli

During all three experimental conditions the same background auditory stimulus was presented. Brainstem responses were elicited in response to the synthesized speech syllable /da/. The /da/ stimulus was derived from a 250ms long 3-formant speech syllable synthesized in Klatt (1980), resampled at 25000 Hz in Praat (Boersma, 2001) to match the sampling rate of the EEG recordings, and shortened in duration using the Lengthen (overlap-add) function in Praat. The resulting stimulus was 80ms long with a 10ms release burst with typical alveolar spectral shape followed by a 10ms voice onset time, a level fundamental frequency (F0: 100Hz), and level first and third formants (F1: 768Hz, F3: 2630Hz, respectively) over the duration of the syllable. The second formant cued the place of articulation varying from 1440Hz from the onset of voicing over a 20ms transition to 1240Hz for the remainder of the syllable duration. Stimuli were presented at a rate of about 7 per second, with a variable inter-stimulus interval of 51-56ms. Within the tapping conditions, the speech syllable was presented in blocks of 1000 repetitions with a 2-minute break between blocks, at which time participants were not required to perform the tapping task.

Procedure

This experiment was conducted in one 90-minute session. At the start of the experimental session, participants were asked to read and sign an informed consent form and then completed a Musical Experience Questionnaire and a demographics questionnaire. Presentation order of the three experimental conditions was completely counterbalanced. Following completion of each experimental condition, participants were asked to rate how difficult they found the task to be on a scale of 1 to 5, with 1 being not difficult at all and 5 being very difficult.

Electrophysiology

Stimulus presentation

All EEG recording and stimulus presentation procedures were conducted in the same manner as in Experiment 1. Background auditory stimuli were presented binaurally via Etymotic Research ER-3A insert tubephones at 65-75dB at the ear. Stimuli were presented at alternating polarities, with each polarity presented a total of 1500 times (3000 total renditions of the stimulus) in each condition of the experiment. For the tapping conditions, the presentation of the stimulus was broken up into three smaller blocks of 1000 repetitions each to give participants a break from the tapping task.

Recording and data processing procedure

All EEG recording procedures were identical to that of Experiment 1. Brainstem responses were collected at a sampling rate of 25kHz using scalp electrodes and a BrainVision actiChamp amplifier system (BrainProducts). Four Ag-AgCl electrodes were placed on the scalp in a vertical montage (Hood, 1998), with the active electrode placed at central vertex (Cz), linked reference electrodes placed on each earlobe and a ground electrode on the forehead. Contact impedances were kept under 5kOhms. Recordings were made in BrainVision PyCorder (BrainProducts) in continuous mode with an online filter of 0.1-3000 Hz. Data from the photocell were also recorded by the actiChamp system at the same sampling rate to monitor the tapping motions of the participants.

All preprocessing of the neurophysiological data was conducted using BrainVision Analyzer 2 software (BrainProducts). For analyses conducted on the brainstem response, recordings were bandpass filtered off-line from 70 to 3000 Hz (12 dB/octave rolloff) using a Butterworth filter. A 60Hz notch filter was also applied to the EEG recordings to eliminate line noise. Each /da/ was epoched with a window of -10 to 120ms. Artifacts were rejected at a +/-35µV voltage threshold. On average, less than 1% of segments were discarded due to artifact. The remaining pool of artifact-free epochs were baseline corrected using an interval of -10 to 0ms, averaged within each condition and visually inspected for the presence of a frequency following response and a high signal to noise ratio. The averaged waveforms were exported for further analysis, converted to wav files in MATLAB (The MathWorks, Natick, MA) and then the phase-locked component of the brainstem response was examined in Praat (Boersma, 2001). Spectral slices were taken of each wav file and the peak in the response spectrum nearest the F0 (100Hz) and F1 (768Hz) of the /da/ stimulus was obtained for each participant (Skoe et al., 2013).

Averaged waveforms for each block within each tapping condition were also extracted to monitor quality control of the FFR throughout the recordings. After using the same filter as described above, recordings for the two tapping conditions were segmented into three separate blocks, one for each block of recording between breaks, and all further preprocessing steps were equivalent to those described above. Repeated measures ANOVAs for each tapping condition revealed that there is no difference in the amount of spectral power of the FFR at 100Hz across blocks in the 40bpm condition ($F(2,46) = .03$, $p < 0.97$) or in the 80bpm condition ($F(2,46) = 0.03$, $p < 0.97$).

Tapping data from the photocell was also processed in the same manner as in Experiment 1 using the BrainVision Analyzer 2 software (BrainProducts). Photocell recordings were high pass filtered at .1Hz (12 dB/octave rolloff) using a Butterworth filter. The onset, peak and offset of each tap was marked based on change in signal amplitude above and below a baseline threshold. Auditory stimulus events were then epoched based on whether one of these markers for tap onset, peak or offset occurred within a window of -10 to 120ms of the onset of the /da/. After artifact rejection, baseline correction and averaging, using the same procedures as described above, similar spectral power analyses were performed in Praat to compare the amount of spectral energy at F0 and F1 in the brainstem response for events in which a tap and /da/ occurred concurrently with events in which a /da/ was played in the absence of a tap.

EEG recordings were processed and analyzed to measure cortical theta, alpha, beta and gamma band power as well. For these cortical EEG analyses, recordings were bandpass filtered off-line from 0.1 to 120Hz (12 dB/octave rolloff) using a Butterworth filter, along with a 60Hz notch filter. After segmenting out the break period between blocks within each tapping condition, recordings were epoched into 1s intervals, an FFT with a Hanning window of 10% length was performed on each 1s epoch to determine frequency power and then those FFTs were averaged for all epochs. Maximum peak spectral power information for each frequency band was extracted from the averaged FFT and compared across conditions.

Due to the small duration of the /da/ stimulus, EEG oscillation segmentation procedures for Taps and No Taps events differed from those performed for the FFR analysis. After inserting markers for the onset, peak and offset of each tap in the photocell recordings, /da/ stimuli were epoched based on whether or not a tap occurred within the window of -10 to 120ms of the onset of the /da/, however, the data were then segmented at a window of -40 to 320ms so that the FFT analysis would be able to measure power within all four frequency bands of interest. This segmentation window led to about 3 /da/ presentations included in each segment on which an FFT was performed. The FFTs of each segment were then averaged over all epochs and maximum peak spectral power information for each frequency band was extracted and compared across Taps and No Taps events.

Results

As stated in the data processing section, the peak in the FFR spectrum nearest the F0 (100Hz) and F1 (768Hz) of the /da/ stimulus was obtained for each participant (Skoe et al., 2013). Mean peak frequency and standard deviation for each condition are presented in Table 3.1.

Condition	F0 (Hz)	F1 (Hz)
Resting	Mean = 96.34 SD = 2.47	Mean = 765.42 SD = 21.10
40bpm	Mean = 97.76 SD = 3.46	Mean = 762.37 SD = 18.30
80bpm	Mean = 97.19 SD = 2.59	Mean = 759.64 SD = 10.12

Table 3.1. Average peak frequency in FFR spectrum for F0 and F1 components of the /da/ syllable for each experimental condition.

Spectral power at the fundamental frequency and first formant of the /da/ stimulus was first compared across the three experimental conditions. A one-way repeated measures ANOVA revealed no significant effect of condition for F0, with mean spectral power at 52.89 dB for the Resting condition (SD = 4.5), 52.19 dB for the 40bpm condition (SD = 5.63) and 52.74 dB for the 80bpm condition (SD = 4.50; Figure 3.1). Once again, individual differences in the pitch tracking of the brainstem response between these three conditions varied greatly, with roughly equal numbers of participants showing enhancement and suppression of the FFR between the Resting condition and the 40bpm condition, the Resting condition and the 80bpm condition and the 40bpm and the 80bpm condition. The analysis of spectral power at the first formant of the /da/, however, revealed a significant main effect of condition ($F(2,46) = 3.47$, $p < 0.05$), with spectral power at F1 significantly smaller in the 80bpm condition (Mean = 16.82 dB, SD = 5.5) compared to the Resting condition (Mean = 20.53 dB, SD = 4.22; $t(23) = -2.26$, $p < 0.05$). There is also a marginally significant difference between the 40bpm condition (Mean = 19.3 dB, SD = 4.87) and the 80bpm condition ($t(23) = -1.94$, $p < 0.06$; Figure 3.2).

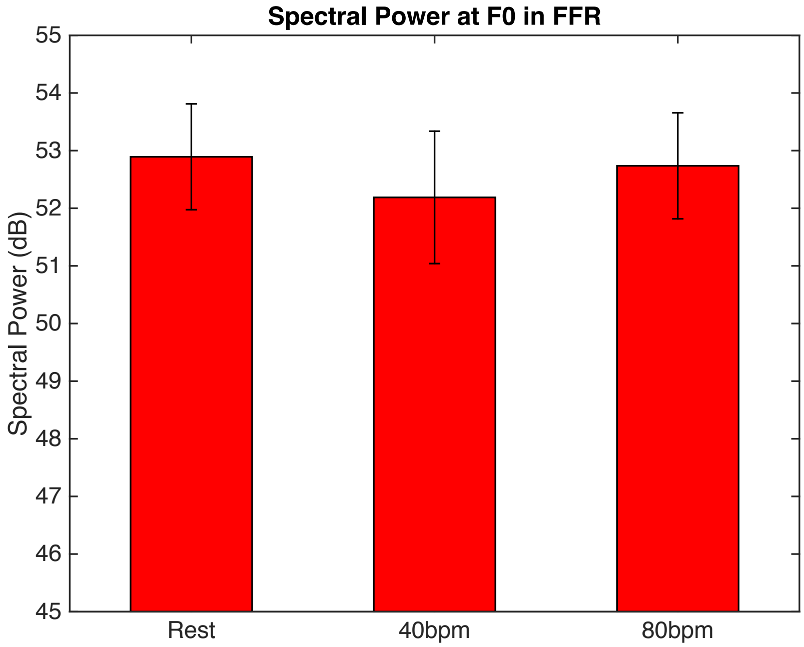

Figure 3.1. Comparison of the FFR for the three experimental conditions. The y-axis plots spectral power at F0 in the FFR in dB.

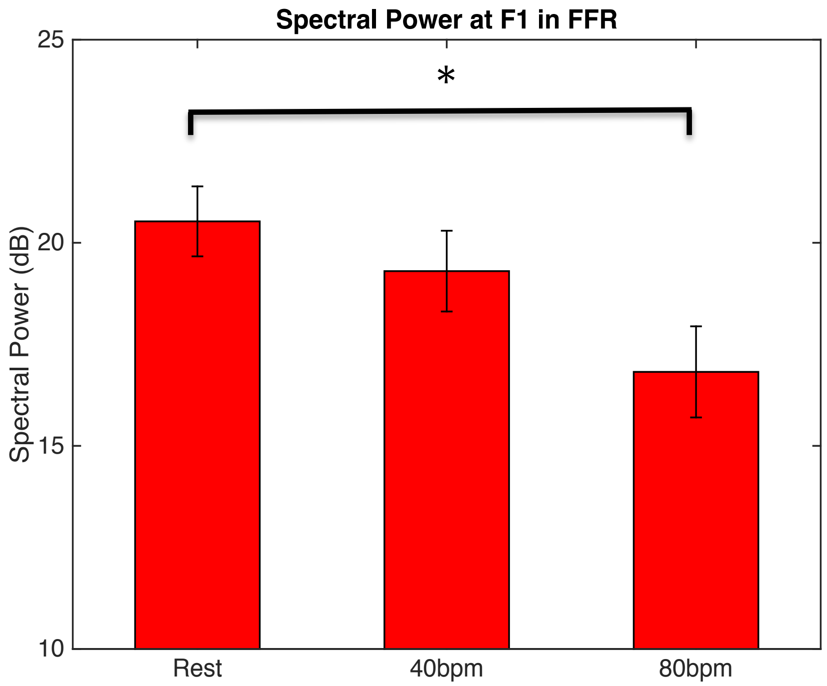

Figure 3.2. Comparison of the spectral power at F1 in the FFR across the three experimental conditions. Asterisk denotes statistical significance (*$p < 0.05$).

Mean peak frequency and standard deviation in the FFR spectrum for each tapping event are presented in Table 3.2.

Condition & Event Type	F0 (Hz)	F1 (Hz)
40bpm Taps	Mean = 97.42 SD = 4.44	Mean = 759.33 SD = 16.22
40bpm No Taps	Mean = 97.9 SD = 3.56	Mean = 748.1 SD = 16.67
80bpm Taps	Mean = 97.96 SD = 7.06	Mean = 766.17 SD = 18.17
80bpm No Taps	Mean = 97.30 SD = 2.51	Mean = 766.79 SD = 19.56

Table 3.2. Average peak frequency in FFR spectrum for F0 and F1 components of the /da/ syllable for each event type within each tapping condition.

Analyses of the FFR for the Taps and No Taps events within the 40bpm and 80bpm conditions also produced significant results. As outlined in the Data Analysis section, /da/ events were segmented based on whether or not they co-occurred with a tap in response to the metronome light. Since the /da/ stimulus was presented at roughly 7 repetitions per second, it took a little over 7 minutes to present 3000 stimulus repetitions in each condition. With a tapping rate of 40 taps were minute in both tapping conditions, this resulted in an average of 323 Taps (SD = 27) events and 2673 No Taps events (SD = 31) in the 40bpm condition and 313 Taps events (SD = 33) and 2676 No Taps events (SD = 32) in the 80bpm condition.

A 2x2 repeated measures ANOVA of spectral power at F0 in the FFR revealed a significant main effect of event type ($F(1,23) = 7.30$, $p < 0.02$; Figure 3.3), with Taps events (Mean = 53.74 dB, SD = 3.92) significantly larger than No Taps events (Mean = 52.68 dB, SD = 4.91). The main effect of condition and the interaction of condition by event type were not significant (p > 0.05). Comparisons of peak spectral power at the F1 of the /da/ also demonstrated clear differences between the Taps events and the No Taps events. A 2x2 repeated measures ANOVA revealed a significant main effect of event type ($F(1,23) = 90.25$, $p < 0.000$), with Taps events (Mean = 31.67 dB, SD = 5.68) significantly larger in peak spectral power than No Taps events (Mean = 28.19 dB, SD = 5.12) as well (Figure 3.4). The main effect of condition and the interaction of condition by event type were not significant ($p > 0.05$).

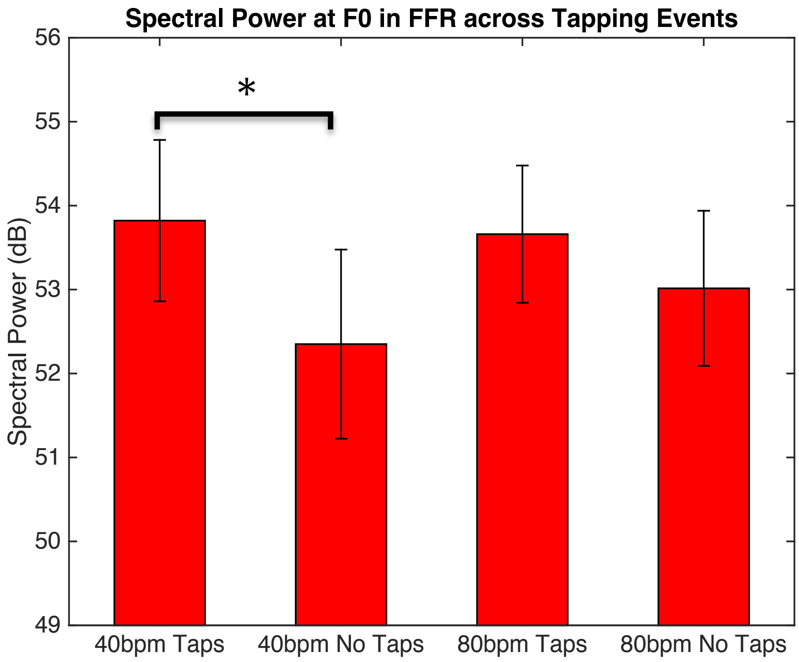

Figure 3.3. Comparison of the spectral power at F0 in the FFR for Taps and No Taps events in the two tapping conditions (40bpm and 80bpm). Asterisk denotes statistical significance ($t(23) = -2.32$, * $p < 0.05$).

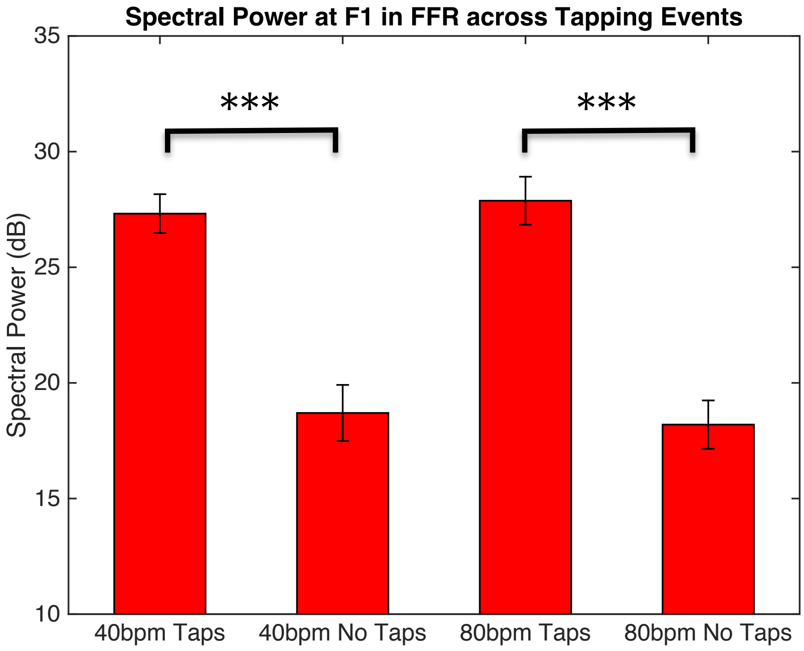

Figure 3.4. Comparison of the spectral power at F1 in the FFR for Taps and No Taps events in the two tapping conditions (40bpm and 80bpm). Asterisk denotes statistical significance (40bpm: $t(23) = -6.45$; 80bpm: $t(23) = -7.4$, *** $p < 0.000$).

As in Experiment 1, cortical oscillation measurements from the EEG recordings were compared across the three experimental conditions. One-way repeated measures ANOVAs of peak spectral power in each of the four frequency bands (theta, alpha, beta and gamma) revealed a significant main effect of condition for theta ($F(2,46) = 6.73$, $p < 0.01$), but no significant effects of condition for the other three oscillation bands (Table 3.3).

Frequency Band	Resting Condition (μV^2)	40bpm Condition (μV^2)	80bpm Condition (μV^2)
Theta (4-8Hz)	Mean = 1.93* SD = 0.79	Mean = 1.82* SD = 0.75	Mean = 1.67* SD = 0.60
Alpha (9-12Hz)	Mean = 1.42 SD = 1.33	Mean = 1.66 SD = 1.33	Mean = 1.49 SD = 0.93
Beta (13-20Hz)	Mean = 0.36 SD = 0.15	Mean = 0.32 SD = 0.15	Mean = 0.33 SD = 0.13
Gamma (25-50Hz)	Mean = .13 SD = 0.08	Mean = 0.11 SD = 0.08	Mean = 0.13 SD = 0.16

Table 3.3. Maximum peak spectral power (μV^2) for theta, alpha, beta and gamma oscillation bands for each experimental condition. Asterisk denotes statistical significance in paired t-tests with other experimental conditions (* $p < 0.05$).

Since the comparison of the spectral power at F0 in the Taps and No Taps events replicated the results of Experiment 1, peak spectral power for each of the four EEG oscillation bands was also compared between the two types of tapping events. For the theta band, a 2x2 repeated measures ANOVA demonstrated significant main effects of condition ($F(1,23) = 5.13$, $p < 0.005$) and event type ($F(1,23) = 7.20$, $p < 0.02$; Figure 3.5), with theta power significantly larger in the 40bpm condition (Mean = 2.70 μV^2, SD = 1.07) than in the 80bpm condition (Mean = 2.46 μV^2, SD = 0.85) and significantly greater for the Taps events (Mean = 2.73 μV^2, SD = 1.05) than for the No Taps events (Mean = 2.44 μV^2, SD = 0.87). The analysis of the gamma band power revealed a significant main effect of event type ($F(1,23) = 7.42$, $p < 0.02$; Figure 3.6) as well, with No Taps events (Mean = 0.11 μV^2, SD = 0.07) containing more gamma power than Taps events (Mean = 0.10 μV^2, SD = .05). Analyses of spectral power in the alpha and beta oscillation bands, however, did not produce significant effects.

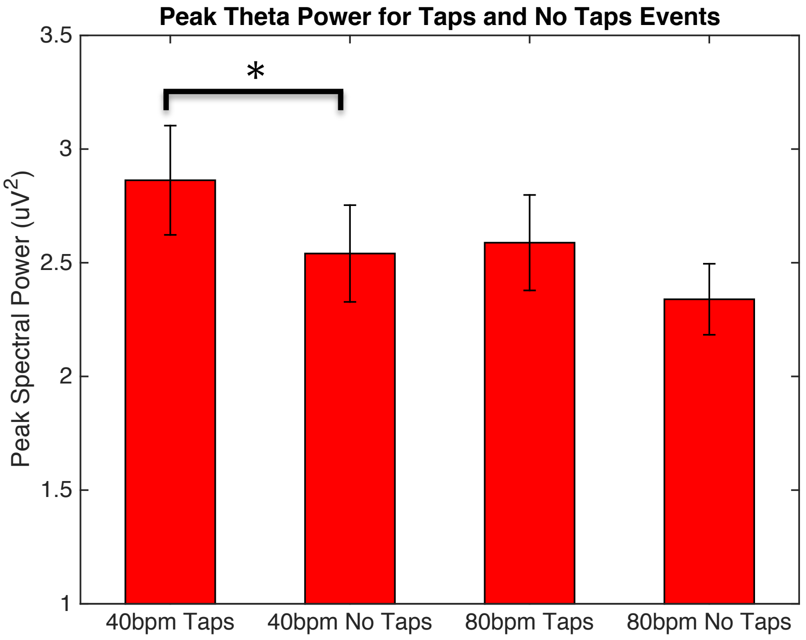

Figure 3.5. Comparison of peak spectral power in the theta frequency band for Taps and No Taps events in the two tapping conditions. The y-axis plots peak spectral power in (μV^2) at 4.58Hz. Asterisks denote statistical significance ($t(23) = -2.55$, * $p < 0.05$).

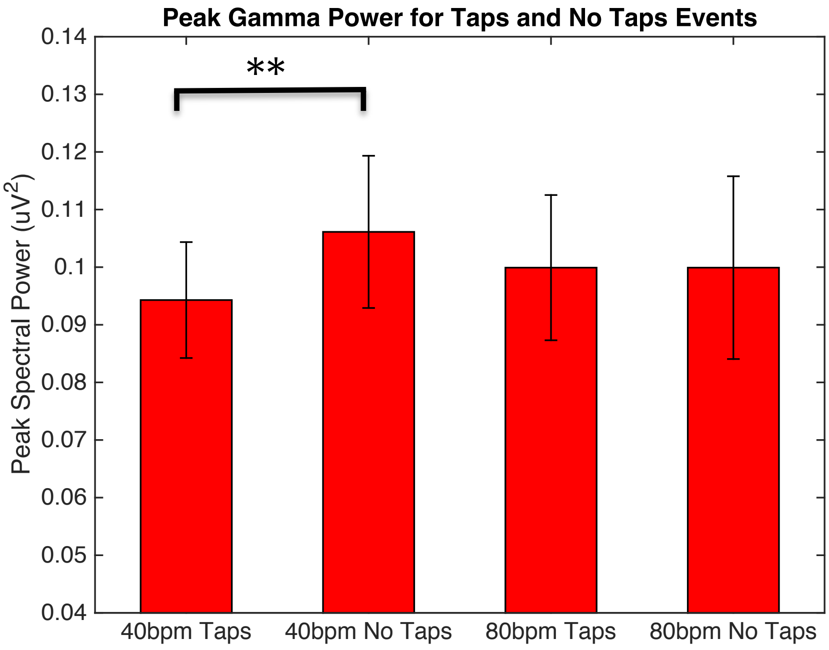

Figure 3.6. Comparison of peak spectral power in the gamma frequency band for Taps and No Taps events in the two tapping conditions. The y-axis plots peak spectral power in (μV^2) at 25.94Hz. Asterisks denote statistical significance ($t(23) = 3.04$, ** $p < 0.01$).

Discussion

The results of this experiment replicate many of the findings of Experiment 1. Once again, there was no overall effect of experimental condition on the representation of F0 in the FFR, with spectral power equal between the Resting, 40bpm and 80bpm conditions. The analysis of spectral F0 power between the Taps and No Taps events also revealed similar results to that of Experiment 1, with Taps events showing an enhancement of the neural representation of pitch information compared to No Taps events. Interestingly, the analyses of the representation of the F1 component of the /da/ syllable produced slightly different results. Spectral power at F1 in the FFR did significantly differ between the three experimental conditions, with the 80bpm condition showing suppression compared to both the Resting condition and the 40bpm condition. The comparison of Taps and No Taps events, however, revealed the same pattern of results as in the F0 analysis, with Taps events significantly larger in spectral power at F1 than No Taps events.

The discrepancy between the F0 and F1 analyses of the FFR across the three experimental conditions is not surprising, as formant information is the acoustic component by which vowels are categorized and pitch information is used to distinguish between talkers. Since there was only one synthetic /da/ presented throughout the entire experiment, which came from one synthetic talker, the listener did not have to discriminate between different talkers within the experiment, which, presumably, diminished the need to augment the amount of neural representation or attention paid to the pitch information (F0) of the syllable between conditions.

A decrease in F1 spectral power between the 80bpm condition and the Resting condition would be considered evidence in support of motor suppression of the auditory brainstem if this result were taken in isolation. However, since a similar decrease was not found between the Resting condition and the 40bpm condition, which contained the same amount and rate of tapping as the 80bpm condition, these results cannot be accounted for by the corollary discharge explanation. A more appropriate explanation of this result would be that the task performed in the 80bpm condition, in which participants were instructed to tap the photocell in response to every other flash of the light and inhibit a tapping response to alternate flashes of the light, required more cognitive resources overall than the tasks performed in the 40bpm and Resting conditions. Since this experiment did not included recordings of the metronome light to determine when the participant was expected to tap the photocell or inhibit a tap, it is still unclear whether the processes involved in inhibiting a motor act are specifically driving the reduction in the representation of F1 in the FFR. However, given that the reduction is only seen for the 80bpm condition, this is likely the case. An additional experiment that recorded or marked instances in which the participant inhibited a motor behavior would be required to fully support this conclusion.

The replication of the pattern of results found in Experiment 1 between the Taps and No Taps events for both the F0 and the F1 components of the syllable are much more conclusive. The enhancement of spectral power in the FFR at F0 for Taps events compared to No Taps events further diminishes the role of corollary discharge projections from the motor cortex on the peripheral auditory pathway. The finding that this enhancement effect can also be seen for the F1 component of the syllable and is significant for both tapping conditions further demonstrates a response-based enhancement of the representation of auditory input in the auditory brainstem and provides additional support for the dynamic attending theory of auditory perception (Large and Jones, 1999). The enhancement of the neural representation of both the F0 and F1 components of a synthetic speech syllable when the motor behavior and the auditory stimulus co-occurred suggests that the tapping behavior induced by the periodic flashing of a metronome light attuned various attention and sensory processes to the time intervals at which perceptual events were expected to occur. Since this temporal tuning led to an enhancement of the perception of the background auditory stimulus, which was irrelevant to the tapping task, these temporally based attention processes are likely involved in multi-modal attention networks that span the entire system, rather than selective networks that inhibit processing in one modality to boost sensory processing in another modality.

Analyses of cortical oscillations from the EEG recordings revealed slightly different results from those found in the first experiment. While Experiment 1 found a significant drop in peak spectral power between the Resting condition and the two tapping conditions for all four frequency bands analyzed (theta, alpha, beta and gamma), significant differences between conditions were only found for theta oscillations. The absence of significant results for the alpha, beta and gamma bands between the three experimental conditions is likely due to the change in procedure for the Resting condition. While in Experiment 1 participants sat quietly and often fell asleep during the 20 minute long Resting condition, in this experiment participants watched a television program for the roughly 7 minute long Resting condition, causing them to stay awake and be much more alert in general than in the first experiment. This meant that general levels of arousal were similar across all three experiment conditions, resulting in equal amounts of spectral power in the alpha, beta and gamma band frequencies. The differences in theta synchrony between the experiment conditions are a bit surprising, given the loss of arousal effects in the other three oscillations bands, but not particularly problematic to the hypotheses of this experiment.

Most importantly, the comparison of power in the theta and gamma bands between the Taps and No Taps events reproduced the results of Experiment 1. Once again, theta power was significantly larger for the Taps events than for the No Taps events, while gamma power was significantly larger for the No Taps events than for the Taps events. This increase in theta band synchrony for Taps events reflects the role of theta activity as a gatekeeper of attention and working memory on an event-by-event basis (Raghavachari et al., 2001). Research has demonstrated that theta activity increases during stimulus presentation and decreases between instances of stimulus perception (Deiber et al., 2007). In this instance, theta band activity likely became synchronized to the flashing metronome light in the tapping conditions, increasing when the light flashed and decreasing between light flashes. Since the flash of the metronome light induced the tapping behavior, the Taps events likely include the presentation of the light as well, reflecting the increase in theta power that occurred in response to the visual stimulus presentation. The reversed direction of the effect in the gamma band further supports the idea that fluctuations in gamma band synchrony reflect preparation or anticipation to engage in another task. While some research has demonstrated that this increase in gamma power, along with an increase in beta power, reflects preparation to perform a motor task (Sanes & Donaghue, 1993), recent findings have demonstrated an increase in gamma power in anticipation of the end of a delay between tasks, suggesting that gamma oscillations may reflect a gating mechanism to accessing working memory (Lundqvist et al., 2016). Since the tapping task in this experiment involves both working memory and motor processes, it is still unclear at this time as to which of these theories best explains the difference in gamma power seen between event types.

The results of this experiment confirm the FFR and cortical EEG results of Experiment 1 using a more ecologically valid auditory stimulus. The enhancement of the phase-locked component of the brainstem response at the fundamental and first formant frequencies of the /da/ syllable provides additional evidence that, rather than interacting with corollary discharge projections from the motor cortex, the auditory brainstem is guided by temporally tuned attention processes that anticipate the time interval in which an event in the environment will occur. The cortical oscillatory results further support this claim by demonstrating that the enhanced FFR is accompanied by an increase in spectral power in cortical oscillations that are associated with increased working memory and attention. However, in order to truly reject the involvement of the motor system in the auditory brainstem response, attentional differences in the frequency following response must be found without a concurrent change in a motor task. A demonstration of changes in the representation of pitch and formant information based solely on attentional differences would provide the appropriate evidence needed to support the role of an attention system in low-level auditory perception.

Chapter 4

Experiment 3: Increased Attentional Demands in an Eriksen-Flanker Task Enhances the Brainstem Frequency Following Response to a /da/ Syllable

Experiments 1 and 2 examined the role of efferent corollary discharge on the responses to acoustic input in the auditory brainstem by inducing a simple motor behavior (i.e., tapping a finger) while listening to sounds. Although cortical auditory responses are suppressed during motor behavior, the results of these experiments exhibited an enhancement of the auditory brainstem response during tapping. Specifically, both experiments found a significant increase in the amplitude of spectral signal information in the brainstem when the presentation of the auditory stimulus coincided with a motor behavior. This suggests that suppressive corollary discharge projections that have been found to inhibit the activity of the auditory cortex (Schneider et al., 2014) do not influence auditory brainstem responses and descending projections in the corticofugal pathway may be controlled by neural processes that regulate attention to incoming signals.

Contrary to the hypotheses for both experiments, neither motor behavior nor putative attention demands in the 80bpm condition suppressed the spectral response to the auditory stimulus in the ABR. This appears to be due to the way in which the visual cue and the tapping motion organized attention to moments in time, which led to an enhancement of the spectral response. However, the difference in demand between the 40bpm and the 80bpm condition (due to the appearance of visual cues that did not result in a tapping behavior in the 80bpm condition) may well not be sufficient to shift attention from the acoustic stimulus. On the other hand, ABR responses may be more automatic than controlled by cognitive processes without extended periods of training or experience and the enhanced spectral response seen in these experiments may be the product of a lower level focusing mechanism produced by the tapping behavior rather than the cognitive networks involved in orienting and focusing attention. To test these alternatives, the flanker task, which is demonstrated to be attentionally demanding (Lavie, 1995), was used to compete with the auditory processing of a sound.

Paradigms that have been used to manipulate attentional demand in examinations of the auditory brainstem response widely vary and produce either inconclusive or conflicting results (Galbraith & Arroyo, 1993; Hairston et al., 2013; Hoormann et al., 1994). Experiment 3 was designed to vary task-based attentional demands by using a modified version of a task that has been well-established as a successful way to manipulate task difficulty through attention demands (see Lavie, 1995; 1997). Participants performed two versions of a modified Eriksen-Flanker task (Eriksen & Eriksen, 1974) while a synthetic /da/ syllable was presented as a background auditory stimulus. The Eriksen-Flanker task requires participants to respond to a target letter that is surrounded by flanker noise letters. In the traditional paradigm, the task consists of two types of trials, trials in which the flanker letters are congruent with the target letter and trials in which the flanker letter is incongruent with the target letter. Responses are faster and more accurate to congruent trials than to incongruent trials (Eriksen and Schultz, 1979). A set of 5 letters, 2 flanker letters on each side of a central target letter, was presented in each trial of the flanker task (e.g. C C P C C). The target letter could be one of four possible letters (C, P, O, R) and the flanker letters were also one of these four letters. The letters C and P corresponded to one target response and the letters O and R responded to the other target response. This meant that congruent trials were instances in which the target letter was surrounded by the flanker letter that corresponded to the same target response and incongruent trials were instances in which the flankers corresponded with the opposite target response.

The ABR to a synthetic /da/ syllable was recorded and the FFR measured while participants engaged in two different versions of the flanker task. In the Easy condition, congruent trials were presented 75% of the time and incongruent trials were presented 25% of the time. In the Hard condition, these percentages were reversed, with incongruent trials making up 75% of the total number of trials and congruent trials appearing 25% of the time. Brainstem responses during these two Flanker conditions were compared to a Resting condition in which participants watched a silent television program or movie with subtitles while the /da/ syllable was presented.

If the descending corticofugal pathway is controlled by efferent corollary discharge from the motor system, then there should be a significant decrease in the phase-locked component of the brainstem response between the Resting condition and the two flanker conditions, but no difference in the FFR between the Easy and the Hard conditions of this experiment, since both conditions require not only the same number of motor movements but a motor behavior in response to every visual stimulus. However, if the descending pathway to the auditory brainstem is modulated by cortical activity that monitors and allocates attention based on situational and behavioral demands, then there should be significant differences in the FFR between the Easy and Hard conditions. If the differences between these conditions show a suppression of the FFR, this would suggest that cross-modal selective attention processes may be controlling efferent projections to the auditory brainstem to inhibit the representation of task-irrelevant acoustic information. An enhancement of the FFR between these two conditions, however, would indicate that attentional processes responsible for the temporal tuning of sensory systems to the environment may be modulating the descending corticofugal pathway.

Materials and Methods

Participants

Recordings from twenty-six adults were included in this experiment (2 participants were eliminated from data analysis due to more than 10% of segments containing artifacts in at least one condition of the experiment). All included participants had normal bilateral audiometric thresholds (air conduction threshold <40dB HL for 500-4000Hz), and no history of learning or neurological disorders. All participants were compensated for their time with either experimental course credit or monetarily.

Behavioral Task

A modified flanker task consisting of incongruent and congruent trials was presented during two of the experimental conditions. On each trial, targets and flankers were selected as one of a set of four possible letters (C, P, O and R). Subjects were told that the letters C and P corresponded with the left hand button press and the letters O and R corresponded with the right hand button press. For congruent trials, when C was presented as a target letter is was flanked by Ps and vice versa (e.g. P P C P P), and when O was presented as a target letter it was flanked by Rs and vice versa. For incongruent trials, if C was presented as a target letter it was flanked by Os (e.g. O O C O O) and vice versa, and if P was presented as a target letter is was flanked by Rs and vice versa. The conditions of the experiment varied by number of congruent and incongruent trials presented. For the Easy condition, congruent trials were presented 75% of the time while the rest of the trials were incongruent. For the Hard condition, 75% of the trials were incongruent and 25% were congruent. In each version of the task, every letter was presented an equal number of times as a target and as flankers.

The flanker task was presented on a laptop with an external keyboard. Labels were placed on the two response keys to indicate the target letters that corresponded to each button. Flanker stimuli were presented in white lettering with a black background in the center of the laptop screen. A white box was also present in the upper left corner of each flanker stimulus for photocell recordings.

Each flanker task began with a set of instructions explaining the task to the participant and instructing the participant to respond to the target letter as quickly as possible while still maintaining accuracy. A practice session consisting of 30 trials, 15 congruent and 15 incongruent, without the simultaneous presentation of the auditory stimulus, was presented at the beginning of each flanker task as well. The test section of the behavioral task consisted of 284 flanker trials to ensure that the duration of the flanker task was at least as long as the duration of the auditory stimulus presentation. Each flanker stimulus was displayed on the screen until the participant pressed one of the two response buttons. Between stimulus presentations a blank screen was presented for a variable duration of 900 to 1100ms. (See Figure 4.1 for illustration of the experimental paradigm)

Background Auditory Stimuli

The same synthetic /da/ syllable from Experiment 2 was used during all three experimental conditions to elicit the brainstem response. The /da/ stimulus was derived from a 250ms long 3-formant speech syllable synthesized in Klatt (1980), resampled at 25000 Hz in Praat (Boersma, 2001) to match the sampling rate of the EEG recordings, and shortened in duration using the Lengthen (overlap-add) function in Praat. The resulting stimulus was 80ms long with a 10ms release burst with typical alveolar spectral shape followed by a 10ms voice onset time, a level fundamental frequency (F0: 100Hz), and level first and third formants (F1:

768Hz, F3: 2630Hz, respectively) over the duration of the syllable. The second formant cued the place of articulation varying from 1440Hz from the onset of voicing over a 20ms transition to 1240Hz for the remainder of the syllable duration. Stimuli were presented at a rate of about 7 per second, with a variable inter-stimulus interval of 51-56ms.

Procedure

This experiment was conducted in one 90-minute session. Presentation order of the three experimental conditions was completely counterbalanced. During the Resting condition, participants sat quietly and watched a television show or movie of their choosing with subtitles while the auditory stimulus was presented. For each of the flanker tasks, participants first read through the instruction screens and performed the practice session without hearing the auditory stimulus. Once the practice session was completed, auditory stimulus presentation began and the participants continued on to the test section of the task. (See Figure 4.1 for illustration of the experimental paradigm.)

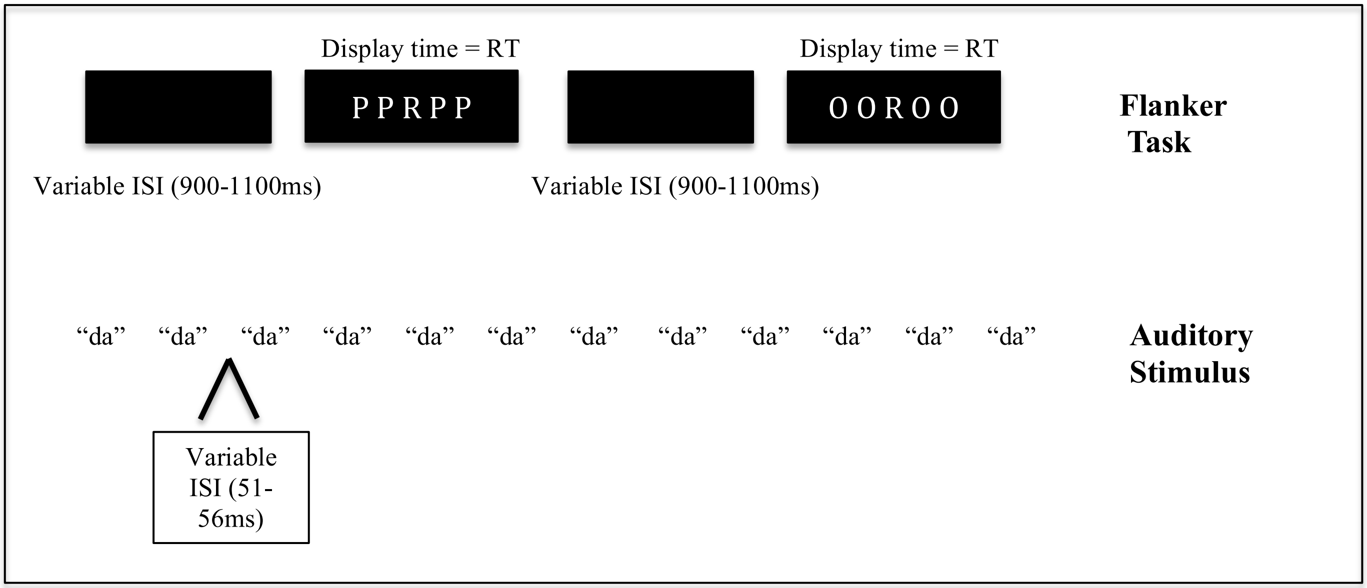

Figure 4.1. Illustration of the flanker task conditions. Within each flanker task, participants responded to two types of trials, congruent trials, in which the target letter and the flanker letters corresponded to the same response button, and incongruent trials, in which the target letter and the flanker letters corresponded to opposing response buttons. For the Easy task condition, 75%

of trials were congruent and 25% were incongruent. For the Hard task condition, 25% of trials were congruent and 75% were incongruent. Throughout both flanker tasks, the /da/ speech syllable was presented at a rate of about 7 repetitions per second.

Electrophysiology

Stimulus presentation

All EEG recordings and stimulus presentation procedures were conducted in the same manner as in Experiments 1 and 2. Background auditory stimuli were presented binaurally via Etymotic Research ER-3A insert tubephones at 65-75dB at the ear. Stimuli were presented at alternating polarities, with each polarity presented a total of 1500 times (3000 total renditions of the stimulus) in each condition of the experiment.

Recordings

All EEG recording procedures were identical to that of Experiments 1 and 2. Brainstem responses were collected at a sampling rate of 25kHz using scalp electrodes and a BrainVision actiChamp amplifier system (BrainProducts). Four Ag-AgCl electrodes were placed on the scalp in a vertical montage (Hood, 1998), with the active electrode placed at central vertex (Cz), linked reference electrodes placed on each earlobe and a ground electrode on the forehead. Contact impedances were kept under 5kOhms. Recordings were made in BrainVision PyCorder (BrainProducts) in continuous mode with an online filter of 0.1-3000 Hz. Data from the photocell were also recorded by the actiChamp system at the same sampling rate to record the occurrence of the presentation of each flanker stimulus.

Data Analysis

Behavioral Task

Performance on the flanker task was evaluated based on response times and accuracy. Response times were measured as the time between the onset of the trial, when the flanker stimulus was presented on the screen, and the time at which the participant responded with a button press. Accuracy was measured as whether the participant responded with the button press that corresponded to the target letter in the trial. Before mean response times were calculated, response times for incorrect trials as well as response times that were more than +/- 2.5 standard deviations from the mean response time for each participant were culled. Mean response times and total number of incorrect trials were then calculated separately for congruent and incongruent trials within each flanker task condition (Easy and Hard).

EEG Preprocessing and Analyses

All preprocessing of the neurophysiological data was conducted using BrainVision Analyzer 2 software (BrainProducts). For analyses conducted on the brainstem response, recordings were bandpass filtered off-line from 70 to 3000 Hz (12 dB/octave rolloff) using a Butterworth filter. A 60Hz notch filter was also applied to the EEG recordings to eliminate line noise. Each /da/ was epoched with a window of -10 to 120ms. Artifacts were rejected at a +/- 35μV voltage threshold. On average, less than 1.5% of segments were discarded due to artifact. The remaining pool of artifact-free epochs were baseline corrected using an interval of -10 to 0ms, averaged within each condition and visually inspected for the presence of a frequency following response and a high signal to noise ratio. The averaged waveforms were exported for further analysis, converted to wav files in MATLAB (The MathWorks, Natick, MA) and then the phase-locked component of the brainstem response was examined in Praat (Boersma, 2001). Spectral slices were taken of each wav file and the peak in the response spectrum nearest the F0 (100Hz) and F1 (768Hz) of the /da/ stimulus was obtained for each participant (Skoe et al., 2013).

Data from the photocell was also processed to examine the brainstem response when flanker stimuli were present on the screen compared to instances in which an auditory stimulus was presented between flanker trials. Using the BrainVision Analyzer 2 software (Brain Products), photocell recordings were high pass filtered at 1 Hz (12 dB/octave rolloff) using a Butterworth filter. The onset of each flanker trial was marked based on change in signal amplitude above and below a baseline threshold. Flanker trials were then epoched from the trial onset marker to 650ms after trial onset, which was the mean response time for the two flanker tasks across all participants. These flanker trial epochs were then further segmented based on the onset of the /da/ with a window of -10 to 120ms. Segments for /da/ stimuli that did not co-occur with a flanker trial were identified as any auditory stimulus presentation that did not have a trial onset marker within the window of -650 to 120ms of the onset of the /da/. After artifact rejection, baseline correction and averaging for both sets of segments, using the same procedures as described above, spectral power analyses were performed in Praat to compare the amount of spectral energy at F0 and F1 in the brainstem response for events in which the flanker stimulus and the /da/ occurred concurrently with events in which a /da/ was played in the absence of a flanker stimulus.

EEG recordings were processed and analyzed to measure cortical theta, alpha, beta and gamma band power as well. For these cortical EEG analyses, recordings were bandpass filtered off-line from 0.1 to 120Hz (12 dB/octave rolloff) using a Butterworth filter, along with a 60Hz notch filter. Recordings were epoched into 1s intervals, an FFT with a Hanning window of 10% length was performed on each 1s epoch to determine frequency power and then those FFTs were averaged for all epochs. Maximum peak spectral power information for each frequency band was extracted from the averaged FFT and compared across conditions.

Results

Behavioral Task

Comparisons of response times as well as accuracy were performed across the two flanker task conditions. A 2x2 repeated measures ANOVA of Condition (Easy vs Hard) by Trial Type (Congruent vs Incongruent) revealed a significant main effect of Trial Type ($F(1,25) = 20.75$, $p < 0.000$) with response times significantly longer for incongruent trials (Mean = 656.52ms, SD = 184.37) than for congruent trials (Mean = 613.16ms, SD = 142.21; Figure 4.2). There was no significant main effect of Condition or significant interaction between Condition and Trial Type ($p > 0.05$). A 2x2 repeated measures ANOVA comparing percent incorrect trials produced a significant main effect of Trial Type ($F(1,25) = 35.85$, $p < 0.000$) as participants committed a higher percentage of response errors for incongruent trials (Mean = 6.79%, SD = 3.62) than for congruent trials (Mean = 3.39%, SD = 2.18). There was no significant main effect of Condition or significant interaction between Condition and Trial Type ($p > 0.05$; Figure 4.3).

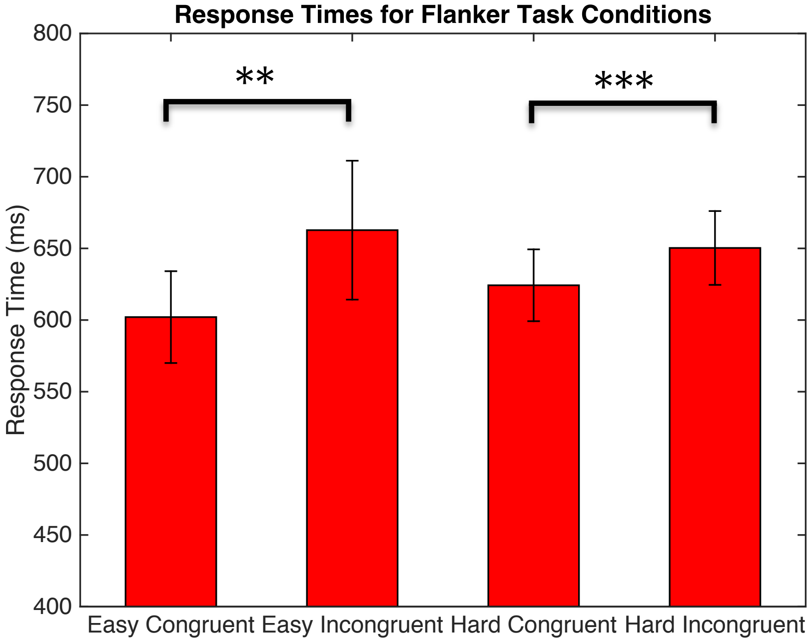

Figure 4.2. Comparison of response times for each trial type across flanker task conditions. Asterisk denotes statistical significance (Easy: $t(25) = 3.4$, ** $p < 0.01$; Hard: $t(25) = 5.59$, *** $p < 0.000$).

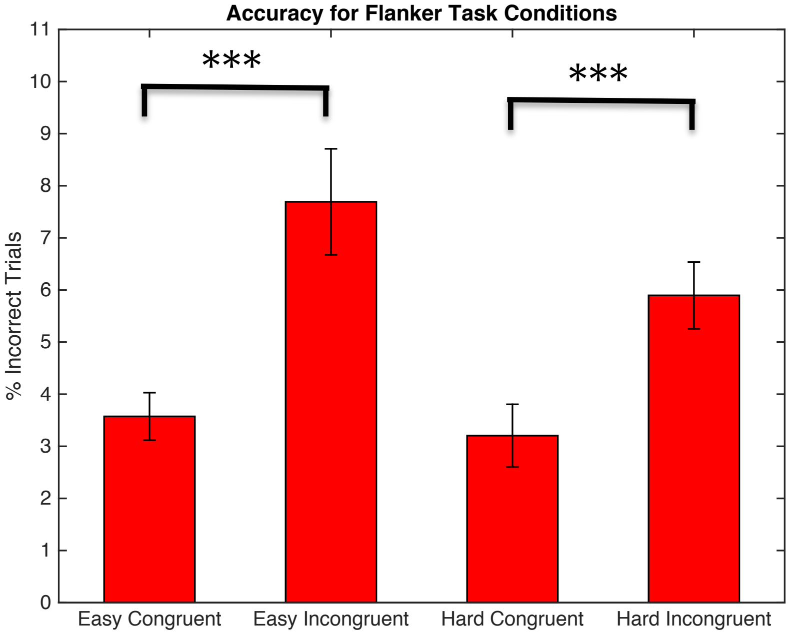

Figure 4.3. Comparison of percent errors for each trial type across flanker task conditions. Asterisk denotes statistical significance (Easy: $t(25) = 5.02$, *** $p < 0.000$; Hard: $t(25) = 4.43$, *** $p < 0.000$).

Frequency Following Response

As stated in the Methods section, the peak in the FFR spectrum nearest the F0 (100Hz) and F1 (768 Hz) of the /da/ stimulus was obtained for each participant (Skoe et al., 2013). Mean peak frequency and standard deviation for each condition are presented in Table 4.1.

Condition	F0 (Hz)	F1 (Hz)
Resting	Mean = 97.75 SD = 2.25	Mean = 765.03 SD = 25.11
Easy	Mean = 97.43 SD = 3.67	Mean = 781.32 SD = 22.12
Hard	Mean = 98.16 SD = 3.44	Mean = 772.78 SD = 23.04

Table 4.1. Average peak frequency in FFR spectrum for F0 and F1 components of the /da/ syllable for each experimental condition.

Spectral power at the fundamental frequency and first formant of the /da/ stimulus was first compared across the three experimental conditions. A one-way repeated measures ANOVA exhibited a significant main effect of condition for F0 ($F(2,50) = 3.88$, $p < 0.05$), with spectral power at F0 significantly larger in the Hard condition (Mean = 53.45 dB, SD = 5.95) compared to the Resting condition (Mean = 51.98 dB, SD = 5.24; Figure 4.4). Analyses of spectral power at the first formant of the /da/ did not produce a significant main effect of condition ($p > 0.05$, Figure 4.5).

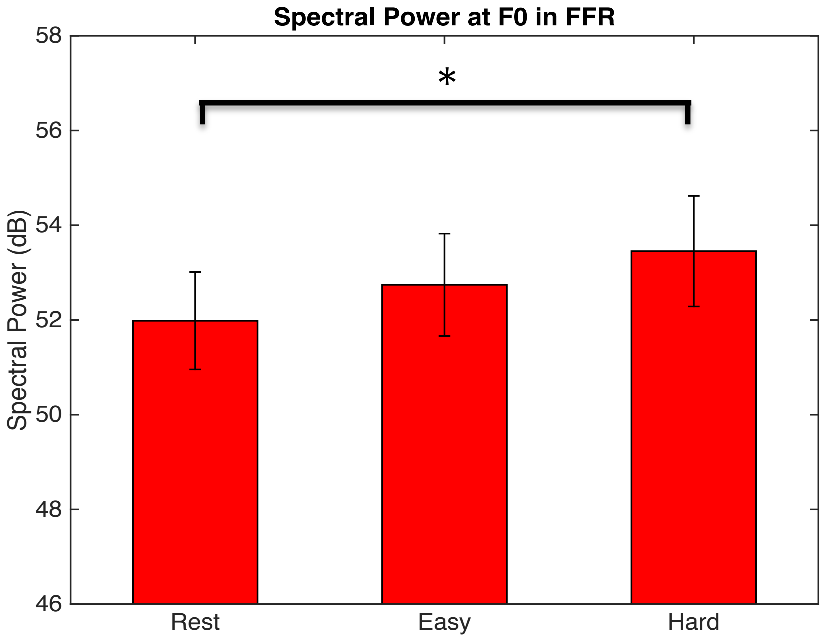

Figure 4.4. Comparison of the spectral power at F0 in the FFR across the three experimental conditions. Asterisk denotes statistical significance ($t(25) = 2.44$, * $p < 0.05$).

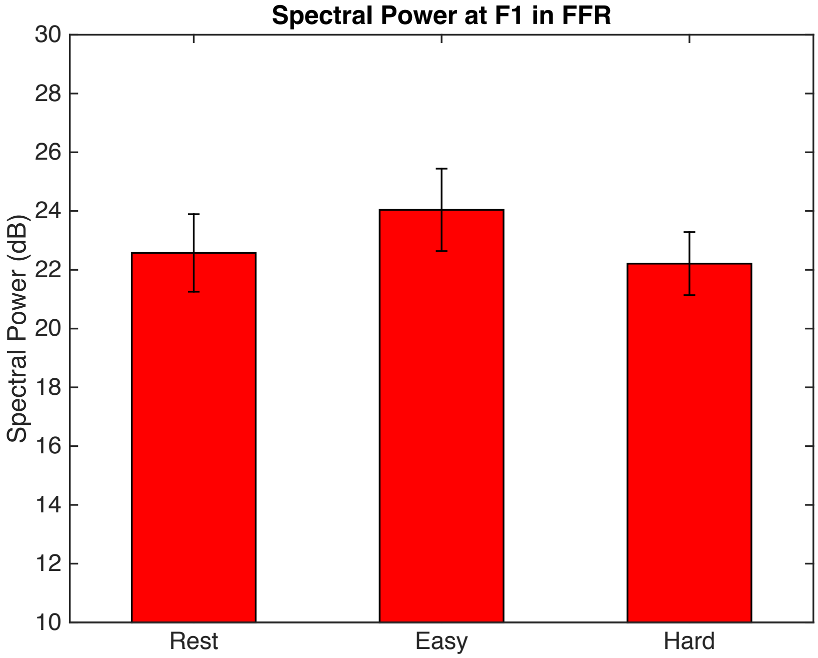

Figure 4.5. Comparison of the spectral power at F1 in the FFR across the three experimental conditions ($F(2,50) = 0.95$, $p < 0.39$).

Mean peak frequency and standard deviation in the FFR spectrum for each flanker task event are presented in Table 4.2.

Condition & Event Type	F0 (Hz)	F1 (Hz)
Easy Flanker	Mean = 99.14 SD = 4.21	Mean = 767.68 SD = 20.64
Easy No Flanker	Mean = 97.17 SD = 4.12	Mean = 769.82 SD = 23.96
Hard Flanker	Mean = 97.90 SD = 3.60	Mean = 759.97 SD = 21.21
Hard No Flanker	Mean = 98.51 SD = 3.62	Mean = 769.36 SD = 24.48

Table 4.2. Average peak frequency in FFR spectrum for F0 and F1 components of the /da/ syllable for each event type within each flanker task condition.

Spectral power in the FFR at F0 and F1 of the /da/ was compared between Flanker and No Flanker events within the Easy and Hard conditions as well. As outlined in the Data Analysis section, /da/ events were segmented based on whether or not they co-occurred with a flanker trial. This resulted in an average of 1004 Flanker events (SD = 56) and 1479 No Flanker events (SD = 99) in the Easy condition and 998 Flanker Events (SD = 55) and 1488 No Flanker Events (SD = 85) in the Hard condition.

A 2x2 repeated measures ANOVA for spectral power at F0 in the FFR did not produce significant main effects of Condition or Event Type or a significant interaction between the two factors ($p > 0.05$; Figure 4.6). Similar analyses of spectral power at F1 in the FFR also did not reveal significant main effects or an interaction between the two factors ($p < 0.05$; Figure 4.7).

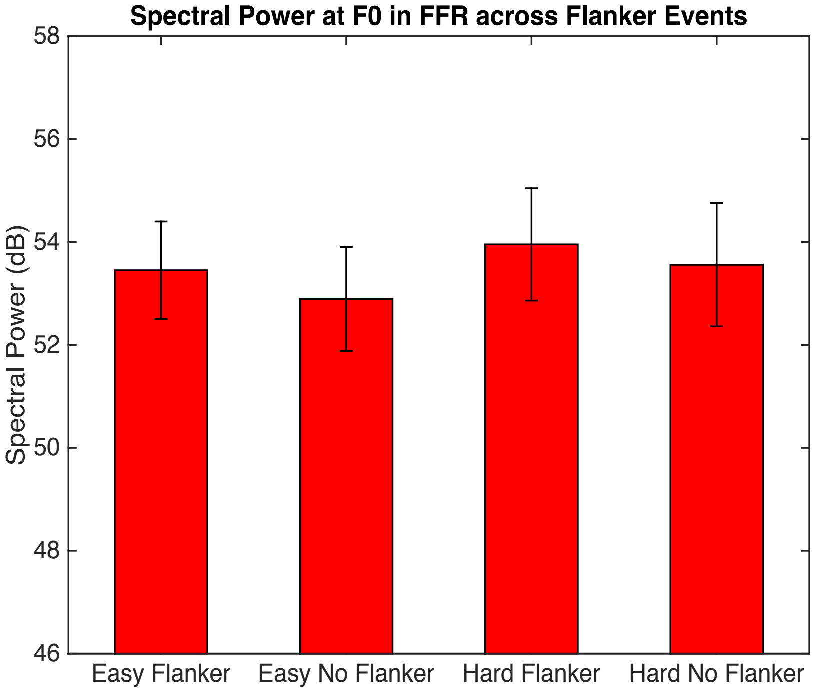

Figure 4.6. Comparison of the spectral power at F0 in the FFR for Flanker and No Flanker events in the flanker conditions (Easy and Hard).

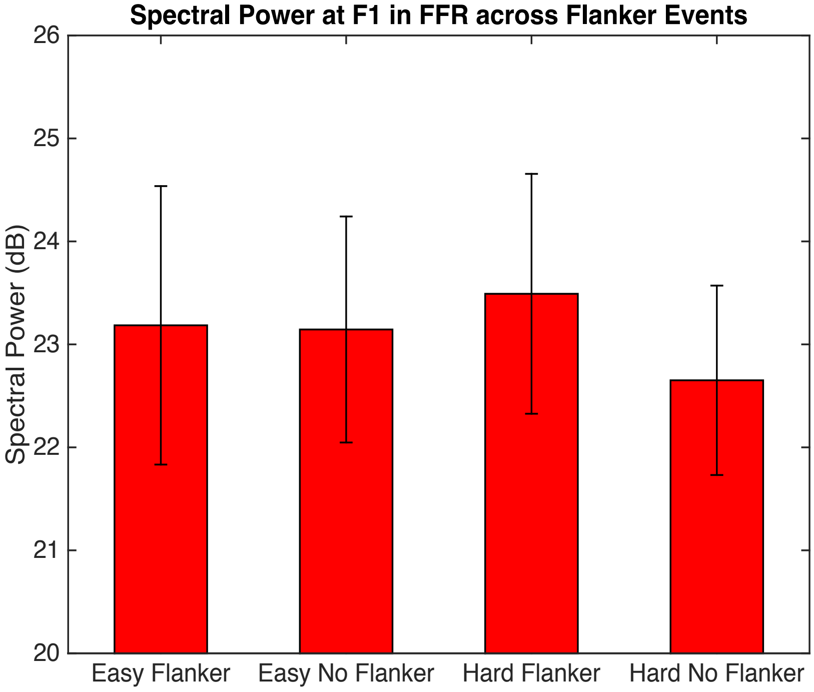

Figure 4.7. Comparison of the spectral power at F1 in the FFR for Flanker and No Flanker events in the flanker conditions (Easy and Hard).

Cortical Oscillations

As in Experiments 1 and 2, cortical oscillation measurements from the EEG recordings were compared across the three experimental conditions. One-way repeated measures ANOVAs of peak spectral power in the four frequency bands (theta, alpha, beta and gamma) revealed a significant main effect of condition for beta ($F(2,50) = 5.27$, $p < 0.01$), but no significant effects of condition for the other three oscillation bands (Table 4.3).

Frequency Band	Resting Condition (μV^2)	Easy Condition (μV^2)	Hard Condition (μV^2)
Theta (4-8Hz)	Mean = 1.79 SD = 0.87	Mean = 2.22 SD = 1.45	Mean = 2.26 SD = 1.33
Alpha (9-12Hz)	Mean = 1.60 SD = 1.63	Mean = 1.64 SD = 1.22	Mean = 1.69 SD = 1.40
Beta (13-20Hz)	Mean = 0.38* SD = 0.19	Mean = 0.50 SD = 0.32	Mean = 0.49 SD = 0.27
Gamma (25-50Hz)	Mean = 0.15 SD = 0.14	Mean = 0.17 SD = 0.20	Mean = 0.14 SD = 0.12

Table 4.3. Maximum peak spectral power (μV^2) for theta, alpha, beta and gamma oscillation bands for each experimental condition. Asterisk denotes statistical significance in paired t-tests with other experimental conditions (* $p < 0.05$).

Discussion

The aim of this experiment was to examine whether task-specific changes in attentional demand could affect the degree to which the auditory brainstem response represents the acoustic information of an auditory stimulus. In order to manipulate attentional demand, two versions of a flanker task were given to participants, an Easy version and a Hard version. Within those two versions of the task, a proportion of the trials were congruent and a proportion of the trials were incongruent. The behavioral analyses of the flanker task revealed that participants were significantly slower in responding to the target letter when presented within an incongruent trial. This effect was consistent across versions of the flanker task, but significant differences in response times were not exhibited between the two difficulty levels of the task.

Analyses of the brainstem responses across the three experimental conditions revealed that the FFR to the F0 of the /da/ stimulus was enhanced for the Hard condition compared to the Resting condition. While the statistical analyses only revealed a significant difference in the FFR between the Resting condition and the Hard condition, a comparison of mean spectral power at F0 for all three conditions demonstrates that the FFR became increasingly stronger as the attentional demands of the task increased, with the mean spectral power at F0 for the Easy condition larger than for the Rest condition and mean spectral power for the Hard condition still larger than the Easy condition.

The lack of significant change in the FFR for the Easy condition compared to the Resting and Hard conditions could be due to the fact that, behaviorally, there was not a significant difference in response time between the Easy condition and the Hard condition, suggesting that the manipulation of difficulty level across conditions was unsuccessful. However, the increase in spectral power at F0 in the FFR results, coupled with the general task requirements of the three experimental conditions indicates that the Hard condition was more difficult than the Easy and Rest conditions. The introduction of a visual task that requires participants to respond as quickly as possible while maintaining accuracy will be more difficult than watching a video with subtitles while listening to a repetitive auditory stimulus, making the Easy condition more difficult than the Resting condition. Given the brainstem response results of Experiments 1 and 2, the increased spectral power at F0 in the FFR for the Hard condition compared to the Rest and Easy conditions suggests that the Hard flanker task required even more attention than the Easy flanker task as well. Though the parameters of the two flanker conditions were not disparate enough to elicit overt behavioral differences that reflect changes in difficulty level, the brainstem response results suggest that the Hard flanker task was indeed more challenging than both the Easy task and the Rest condition, and, thus, required additional attention resources.

The significant increase in the FFR between the Resting condition and the Hard condition and the trend of increasing spectral power across all three conditions provide further evidence against the influence of efferent corollary discharge from the motor system on the activity of the auditory brainstem, since the introduction of the flanker task caused an increase in the representation of pitch information in the FFR, rather than a decrease, as the corollary discharge explanation would predict (Blakemore et al., 2000). More importantly, it provides some of the first evidence for an enhancement of the auditory brainstem response during a visual attention task. Several researchers have attempted to determine the role of attention demands on the auditory brainstem responses, with varying and somewhat conflicting results. For example, while Galbraith and colleagues (2003) determined that FFRs to auditory stimuli are larger when participants perform an auditory discrimination task than when they are engaged in a visual discrimination task, other research has demonstrated that, compared to a resting condition in which auditory stimuli are presented without any simultaneous task requirement, the pitch strength of the FFRs to tones significantly decreased when either an auditory or a visual task was performed (Hairston et al., 2013). Even investigations of the effect of selective auditory attention on the auditory brainstem response have not produced conclusive results. When attention is directed to one of two dichotic pure tones, the effect of endogenous allocation of spatial selective attention on the FFR is unclear, as differences in the FFR to the attended tone compared to the unattended tone are inconsistent within recording sessions (Galbraith & Arroyo, 1993). However, directing auditory attention to one of two dichotic vowels, rather than pure tones, leads to a larger FFR response to the fundamental frequency of the attended vowel compared to the unattended vowel (Galbraith et al., 1998). This experiment provides some of the first evidence that the spectral power of the auditory brainstem response can be increased by the allocation of selective attention to a visual task, suggesting that increased attentional demands in any modality may boost the low-level processing of all sensory input.

The results of the cortical EEG oscillation analyses differed from those of Experiments 1 and 2. Here, the comparison of spectral power across the three experimental conditions revealed only a significant difference within the beta frequency band, with beta power significantly increased for the flanker conditions compared to the Resting condition. This increase in beta power during the flanker task could be due to the motor response that terminated each flanker trial, as research has exhibited increases in beta power, which are coupled to voluntary motor contractions (Conway et al., 1995). On the other hand, the increase in beta power could reflect the increased amount of attention needed to perform the flanker task, as several researchers have demonstrated a top-down effect of beta rhythms on the bottom-up transmission of visual input in primary visual cortex based on selective attention processes (Lee, Whittington & Kopell, 2013). Cortical EEG measurements across the entire scalp, however, would have to be collected in order to determine which of these two explanations best explains the data.

The comparison of events in which the presentation of a flanker stimulus and the presentation of the /da/ stimulus co-occurred with instances in which the /da/ stimulus was presented alone did not reveal any differences in the FFR to the F0 or the F1 of the /da/. However, a difference between event types within the flanker task conditions might be clearer with a different analysis process. In the current experiment, flanker trials were segmented from the moment at which the flanker stimulus was presented until 650ms after flanker stimulus onset, as 650ms was roughly the average response time for all trials within both flanker conditions across all participants. The variance of response times across participants, however, indicates that 650ms may not be the most accurate estimation of response time for certain participants or even for all trials within a participant, confounding the factor of event type, since some Flanker segments likely included flanker trials that ended earlier than 650ms after stimulus onset and other Flanker segments probably failed to include the entire flanker trial. The comparison of flanker event types would likely be better suited to a segmentation process in which the response time of each trial within each participant is used to determine segment length for each flanker trial. This form of segmentation would ensure that the duration of each flanker segment does not exceed or cut short the duration of each flanker trial and thus keep each type of flanker event epoched appropriately. It is possible that the current analyses comparing Flanker to No Flanker events did not yield significant results for the FFR measurements because the way in which these event types were segmented did not adequately separate the attention processes involved in completing the flanker task from moments in which the auditory stimulus was presented on its own.

The results of this experiment provide some of the first evidence that increased visual attention demands can provoke an enhancement of the processing of auditory input in the auditory brainstem. By increasing the amount of attention required to perform a visual task, with the introduction of a speeded classification task in which participants make errors overall requiring more attention than a resting condition, the amount of spectral power within the auditory brainstem response allocated to the pitch of the auditory stimulus was enhanced as well. These results suggest that an attention system, responsible for monitoring environmental and task demands, may be modulating the degree to which the auditory brainstem represents spectral information from acoustic input. More specifically, the results suggest that, rather than suppressing the response elicited in the brainstem by task-irrelevant auditory stimuli, this attention system may engage in a process that enhances the transmission of all external input when a particular sensory task demands additional attentional resources.

Chapter 5: General Discussion

Most neural models of human auditory and speech perception focus on interactions among cortical structures (Hickok and Poeppel, 2007; Rauschecker & Scott, 2009; Recanzone & Sutter, 2008) such as the relative functional roles of different parts of the superior temporal gyrus (e.g., the lateral belt and the parabelt) and parts of the inferior frontal gyrus (e.g., pars triangularis), but few theories of speech perception, in particular, consider the functional role of interactions between ascending and descending projections between cortex and the brainstem. On the other hand, some research that has focused on auditory brainstem responses and the descending auditory pathway, have hypothesized that learning and attention (Krishnan, 2002; Wong et al., 2007) could modify the ascending patterns of neural activity from acoustic stimulation. The functional role of learning (e.g., Krishnan et al., 2005) is putatively to shape the acoustic prominence of certain auditory information based on experience of what is important for certain kinds of perceptual tasks. While some of the work on the phasic responses in the descending auditory pathway has suggested an important hearing-protection mechanism for noise isolation (Guinan, 2006; Brown, 2011), a different functional role has been hypothesized for attentional control of early auditory processing (de Boer & Thornton, 2008). Just as learning can shape the long-term habitual tonic auditory sensitivities of the ascending sensory system, attention can, in the moment of demand or expectation, change the processing of auditory information to increase or decrease the salience of relevant or irrelevant information. The availability of descending projections, even out to the hair cells in the cochlea suggests that such a system could have important phasic control consequences for dynamic deployment of attention to auditory stimuli and scenes. However, the evidence for such a dynamic attention controlled auditory system has been suggestive but somewhat elusive (Galbraith & Arroyo, 1993;

Hoormann et al., 2004; Hairston et al., 2013). It is important to note that the strongest evidence of phasic control of auditory responses comes primarily from research showing noise protective responses that could be mediated solely in the peripheral auditory system. The present experiments begin to provide stronger evidence in support of this kind of model.

At the outset of the present research two broad questions were considered: (1) Given accounts providing evidence for suppression of cortical auditory responses during motor behavior, is it possible that the descending pathway plays a role in attenuating auditory responses even earlier than the primary auditory cortex, such as in the auditory brainstem? (2) Given that the evidence for motor suppression of cortical auditory responses could be due to divided attention with a focus on the motor behavior, could auditory responses in the brainstem be modulated under phasic attentional control? Thus, the three experiments presented here were designed to examine the influence of efferent projections in the corticofugal pathway on the neural activity responding to acoustic input in the human auditory brainstem. Prior research concerning top-down cortical effects on the activity of the auditory brainstem has often been dominated by the consequences of long-term training and experience with certain types of sounds on the auditory brainstem response (Krishnan et al., 2005; Musacchia et al., 2007; Wong et al., 2007), making it unclear how contextual factors can induce phasic changes in the transmission of acoustic input. When real-time effects of top-down cortical projections on the auditory brainstem response have been examined, the results have often been inconclusive (Galbraith & Arroyo, 1993) or have directly conflicted with the findings of other experiments (Hairston et al., 2013; Hoormann et al., 1994; 2004). Experiments 1 and 2 of this dissertation addressed this issue by investigating whether the initiation of a motor behavior during auditory perception could suppress the neural activity responding to acoustic input in the auditory brainstem, as prior research has demonstrated a suppression of the auditory cortex during movement (Houde et al., 2002). It is important to note that both of these experiments were designed to separately assess motor suppression from attentional suppression. The basic hypothesis was that motor behavior should suppress auditory brainstem responses in much the same way as auditory responses in the primary auditory cortex are suppressed by motor behavior (Schneider et al., 2014). However if motor behavior is held constant between two conditions, but the attention demanded in one condition is greater than in another, this should allow separate assessment of any attention effects. The hypothesis was that increasing attention demand while holding motor behavior constant should suppress auditory brainstem responses more than motor behavior alone, essentially under the premise of divided attention between the auditory and motor tasks (Johnson & Zatorre, 2006). Experiment 3 then examined whether the changes exhibited by the brainstem response during a motor behavior could also be demonstrated with the introduction of a visual attention task to determine the possible cortical origins of the descending projections in the auditory corticofugal pathway.

5.1 Overview of Findings

Taken together, the results of the three experiments presented here demonstrate that the descending corticofugal projections to the auditory brainstem are not controlled simply by inhibitory corollary discharge projections from the motor network. In fact, Experiment 1 revealed that the phase-locked component of the auditory brainstem response elicited by a pure tone is actually enhanced when a motor behavior is enacted. The frequency following response showed increased amplitude at the stimulus frequency thereby increasing the signal-to-noise ratio when the acoustic stimulus co-occurred with a finger tap than when the tone stimulus was presented alone. Experiment 2 replicated these findings by demonstrating that the brainstem response elicited by a more complex stimulus, a very brief synthetic speech syllable increased the spectral response to the pitch and first formant of the syllable when it was heard at the same time that a tapping behavior was executed compared to when it was heard in isolation. These results directly conflict with research demonstrating suppression of the cortical auditory system during motor behavior (Schneider et al., 2014), thus rejecting the hypothesis that corollary discharge from the motor network inhibits the auditory brainstem response via efferent projections in the descending auditory pathway.

Furthermore, the results of Experiments 1 and 2 did not provide any evidence for suppression of the auditory brainstem response due to increased attentional demands either. In both experiments, a second tapping condition was introduced that required participants to discriminate between visual signals that were cues for tapping and those that were not, in an attempt to create greater attention demand for the tapping task and potentially remove attention from the auditory system, leading to a suppression of the auditory brainstem response. However in this condition as well, auditory responses were enhanced during tapping with no clear evidence of divided attention. These results reject the hypothesis that increasing attentional demands on a motor task, by requiring participants to execute or suppress a tapping behavior in response to a visual cue, will result in a suppression of the auditory brainstem response. Since the auditory stimulus in each of these experiments was irrelevant to the motor task, it is possible that with increasing attentional demands in the motor task, the processing of irrelevant sensory input would be diminished. Experiments 1 and 2, however, did not exhibit such an effect of attentional demand, suggesting that the efferent corticofugal pathway may not control auditory processing in the brainstem based on divided attention between multiple modalities.

To directly test the effect of attention on the auditory brainstem response while holding motor responses constant, Experiment 3 gave participants two versions of a visual attention task that varied in level of difficulty, rather than a motor task, while a synthetic speech syllable was presented. The results exhibited an enhancement effect of the frequency following response in the ABR based on directed attention, similar to that found in Experiments 1 and 2. Compared to a baseline rest condition, a more attention demanding (harder) task enhanced the response to the pitch of the speech in just the same way as tapping increased the frequency following response to acoustic stimuli. Moreover, the easier attention task produced a spectral enhancement of syllable pitch in the brainstem that was between the baseline rest condition and the hard condition, demonstrating a graded increase in the enhancement of the brainstem response based on the amount of attention required to perform the task. Combined with the findings of the first two presented experiments, these results suggest that the descending corticofugal pathway may be modulated by an attention network that calculates task-based demands and distributes increasing amounts of attention to all sensory processes when the events of the task at hand require it.

The corticofugal network, however, does not contain only descending projections from higher cortical and subcortical structures to lower anatomical regions in the peripheral auditory pathway. It is the interaction of efferent and afferent projections that are important to understanding these phasic changes in the perception of environmental input. Thus, it is important to place this enhancement effect of the auditory brainstem response within the context of cortical changes in neural activity during the tasks of these experiments as well. To that end, EEG recordings in each experiment were processed and analyzed to examine changes in cortical oscillatory behavior during the same experimental tasks.

The hypothesis that corollary discharge from the motor system will suppress the auditory brainstem response to acoustic input in the same way that it does in primary auditory cortex corresponds with the cortical hypothesis that introducing a tapping task will result in a decrease in alpha band power, as motor behavior induces a decrease in alpha synchrony within the motor system, referred to as mu suppression (Chatrian et al., 1959). If a tapping motion leads to a decrease in the spectral power representing an auditory stimulus in the auditory brainstem response, one would expect to see a corresponding decrease in alpha power during the tapping task representing the execution of the motor behavior. However, only Experiment 1 demonstrated a significant decrease in alpha power during the tapping tasks compared to the baseline rest condition, which was most likely the result of participants falling asleep during the rest condition of this experiment than the introduction of the tapping task. Once participants were given a task that prevented them from falling asleep during the rest condition in Experiment 2, this reduction in alpha synchrony for the tapping task was no longer present. These results reinforce the finding that the descending peripheral auditory pathway is not simply controlled by efferent corollary discharge from the motor system, as neither the auditory brainstem response nor cortical alpha power were suppressed with the introduction of a motor behavior.

A reduction of the neural activity in the brainstem in response to acoustic input on the basis of attention demands holds a corresponding cortical prediction as well. If increased task demands in either the tapping condition or the visual attention task condition lead to a suppression of the auditory brainstem response, as resources are redistributed to increase attention in the modalities that are needed to complete the behavioral task, there should also be changes in cortical theta and alpha band synchrony that reflects this change in attention distribution. Prior research has demonstrated an alpha-mediated attention gating mechanism in which alpha power increases or decreases based on the need to inhibited unnecessary or distractor input (Foxe & Snyder, 2011). Additionally, theta oscillation activity has been hypothesized to be responsible for the temporal organization of working memory information (Roux & Uhlhaas, 2014), with theta power increasing when tasks become more difficult (Sauseng et al., 2005). Based on these findings, a suppression of the auditory brainstem response during increased attentional task demands, as reflected in the (harder) second tapping condition in Experiments 1 and 2 and the harder visual attention task in Experiment 3, should be accompanied by an increase in alpha and theta power during these conditions, reflecting both the increase in the difficulty of the task and the need to suppress the processing of irrelevant auditory information while completing the behavioral task.

In all three presented experiments, however, significant increases in alpha and theta band power were not found between the two task conditions. Instead, Experiment 1 revealed a significant increase in theta and alpha synchrony specifically during tapping events and Experiment 2 showed similar results in theta oscillation activity, just as analyses of the brainstem response exhibited enhancement of the frequency following response to acoustic input during tapping events. Furthermore, Experiment 3 exhibited a significant increase in beta band synchrony for the two visual task conditions compared to the baseline rest condition, just as the representation of the pitch information in the auditory stimulus was enhanced in the brainstem response for these two conditions. While this result was not predicted, it may also reflect the activity of an attention network attuned to the demands of a behavioral task, as prior research has established a selective attention based top-down effect of beta rhythms on bottom-up transmission of sensory input (Lee, Whittington & Kopell, 2013). Taken together, these fluctuations in cortical oscillations and the enhancement of the frequency following response in the auditory brainstem in response to either specific events within the task or to a general increase in task demands suggest that, rather than suppressing unnecessary sensory input when a behavioral task becomes more difficult, the corticofugal pathway may facilitate the interactions between early sensory processing regions and cortical networks responsible for allocating attention to sensory networks at specific moments based on behavioral task demands.

Overall, the hypotheses for the three experiments presented here were each rejected by the brainstem and cortical EEG findings. In the first two experiments, we hypothesized that a motor behavior would suppress the brainstem response to auditory stimuli, yet the frequency following response in the auditory brainstem and cortical oscillations that are usually suppressed when movements occur were enhanced during specific events within the task. We also hypothesized that an increase in attention demand would further diminish the neural representation of irrelevant auditory information in the auditory brainstem during movement, but found an increase in the brainstem response to auditory input as well as an increase in cortical activity associated with attention and working memory processes during tapping events regardless of overall attentional task demands as well. In the third experiment, we predicted that increasing the difficulty level of a visual attention task would diminish the transmission of acoustic input by the auditory brainstem of an acoustic stimulus that was unnecessary to complete the task. However, once again, the results exhibited an increase in spectral power of the frequency following response to the pitch of the auditory stimulus as the task became more difficult.

Taken together, these results indicate that the descending corticofugal pathway is not controlled by cortical networks that simply suppress the processing and transmission of acoustic input when additional behavioral tasks are introduced, but rather may be a part of a widespread system that directs attention to sensory input on an event-related basis. Such a system would not be responsible for weighing the importance of one sensory input over another to enhance relevant information and suppress irrelevant inputs. Instead, this system would monitor overall environmental and behavioral events and increase overall sensory attention to specific moments or events. The following sections discuss the implications of such a system for theories of perception at subcortical and cortical levels. The importance of continuing to establish the interaction between cortical networks and early sensory encoding and transmission within the corticofugal system is discussed as well.

The three experiments presented here each demonstrate an enhancement of early auditory processing at specific moments within a non-auditory task, indicating that attention is modulated by specific task events and allocates additional resources to all environmental input at these particular moments. Previous studies of the auditory brainstem response have demonstrated local enhancement of the frequency following response, but only within the context of an auditory task. For example, Skoe & Kraus (2010) reported an enhancement of the FFR to a repeating note within a melody, but only after the melody was repeatedly presented for 90 minutes without requiring the participants to perform a simultaneous task, indicating more of an attention effect on the brainstem response based on training or experience in the auditory realm than the attentional demands of some behavioral task. Chandrasekaran and colleagues (2009) did demonstrate an event-related enhancement of the FFR, with brainstem responses elicited by a synthetic speech syllable exhibiting enhanced representation of the second and fourth harmonics of F0 of the syllable when it was presented within a repetitive block compared to instances in which the same syllable was presented within a variable block of speech sounds, but the researchers did not include a separate task for the listeners to perform and did not require a response of any kind from the participants, making it unclear whether this enhancement effect was local to the auditory system or was part of a system-wide attention network.

The brainstem findings for Experiments 1 and 2, however, are some of the first to demonstrate an enhancement of the FFR on a moment-to-moment basis that is driven by an attention network attuned to the demands of all environmental and behavioral demands. The comparison of taps and no taps events within each tapping condition of these experiments shows that even within a sequence of events, the auditory stimuli that occur with tapping have enhanced spectral response compared to interleaved auditory stimuli without concurrent tapping. Additionally, the results of Experiment 3 demonstrate that the amount of spectral power dedicated to the pitch information of an auditory stimulus in the auditory brainstem is increased with increasing demands from a separate visual attention task. These results not only demonstrate an event-related enhancement of the FFR, but do so within the context of a behavioral task that does not require the processing of auditory input, suggesting that early transmission of sensory input is modulated in real time by the demands of the task at hand, with all sensory input gaining additional attentional resources when a behavioral task requires them. Moreover, the results of these experiments indicate that the corticofugal system, rather than utilizing the typical feedback and feedforward pathways outlined in most neural models of auditory perception, may rely on the synchrony between internally generated contextual and response templates and the encoding and transmission of environmental input at specific moments or events. Several neural models of attention and working memory support this theory and provide substantial evidence for the role of such a system in perception.

For example, research on attentional control of sensory processing in non-human animals has examined active sensing, the acquisition of sensory inputs via overt motor sampling behaviors, such as whisking or sniffing in animals (Kleinfeld, Ahissar & Diamond, 2006), as a means by which attention may modulate sensory perception. Morillon and colleagues (2015) posit that sensory perception is embedded in the temporal oscillations of the motor system, but modulated by dynamic shifts in attention. Active sensing, which in humans can be attributed to hand or eye movements, creates an oscillatory pattern of sampling sensory inputs through perceptual networks. Peaks in these oscillations allow for optimal processing of sensory inputs while troughs result in minimal amounts of sensation and perception. Morillon and colleagues argue that a dynamic top-down control system monitors this pattern of sensory scanning and utilizes it to optimize sensory processing at predicted or expected points in time. While the results of this dissertation argue against an inhibitory effect of the motor system on the descending corticofugal pathway, the type of top-down control system proposed by these researchers is supported by the results of the first two experiments presented. The periodic motor behavior performed during the presentation of the auditory stimulus may have engaged a system that enhances the encoding and transmission of all sensory input at specific moments in time based on the execution of the finger movement, resulting in an enhancement of the frequency following response to the acoustic input when it coincided with a peak in the oscillatory pattern of the motor driven sensory network.

However, since the motor behavior elicited in the first two experiments was cued by a visual stimulus as well, it is unclear whether the increase in sensory processing at certain points in time was due to the motor movement itself or to the periodic presentation of the visual cue that prompted the motor behavior. Moreover, the results of Experiment 3 demonstrate an enhancement of sensory processing without a rhythmic motor response or the periodic presentation of a visual cue, indicating that the increase in processing of sensory input during these tasks was driven by increased attention to specific task events or moments, not the regular or predictive timing of these events.

On the other hand, rather than a model for a hierarchical system of top-down control in which cortical regions and networks exert control over subcortical structures that initially encode and transmit input, a dynamicist model of top-down control in which perception is achieved through a process of large scale dynamics or oscillations influencing local neuronal behavior by 'enslaving local processing elements' may be more appropriate (Kay & Freeman, 1998; Engel, Fries & Singer, 2001). Such a model not only explains the event-related differences seen in the frequency following response in the presented experiments, but also describes the potential relationship between this early sensory enhancement effect and the corresponding cortical oscillatory changes seen in these experiments. A dynamicist view of top-down neural control proposes that the dynamic synchrony of frequency band oscillations across a large group or multiple groups of neurons can influence smaller neuronal groups by pushing them into the same neural oscillation frequency. This mechanism is not specific to a certain neural area or network, as it can happen across anatomical structures or within a neural region through local lateral inhibition. Applying this model to perception, researchers have hypothesized that top-down effects can be manifested in expectations, which are expressed by the temporal structure of activity patterns prior to stimulus onset. This predictive activity has been measured in association cortices in prefrontal, posterior parietal and inferotemporal cortices as well as subcortical structures. Through a form of temporal binding, the dynamic activity pattern of the large neuronal group influences smaller groups of neurons that encode and transmit input through either top-down projections from cortex to subcortical sensory regions or through lateral inhibition within the subcortical structures themselves. This dynamic model of top-down control of perception may be best suited to explaining the parallels seen between the oscillatory activity at a cortical level and the phase-locked component of the auditory brainstem response in the experiments presented here. It is possible that the dynamic activity pattern of cortical networks involved in attention and working memory, as evidenced by power fluctuations in alpha, theta and/or beta frequency bands, modulated local neural activity throughout the descending auditory pathway, resulting in amplification of early neural transmission of acoustic input at specific moments or events within the behavioral task. However, before this model of cortico-cortical and corticofugal processes can be applied to the peripheral auditory pathway, future research must demonstrate a direct correlation between cortical oscillatory patterns and the activity of the auditory brainstem.

5.2 Future Directions

The results of the experiments presented in this dissertation directly contrast with findings in human and non-human animal research demonstrating a suppression of the cortical auditory system during motor behaviors (Schneider et al., 2014; Houde et al., 2002). Rather than demonstrating a decrease in the representation of auditory input during motor behaviors, the experiments presented reveal fluctuations in the frequency following response related to the events of the behavioral task performed by the participants. Given that all three experiments involved the execution of a motor behavior in response to a visual cue, it is unclear whether the effects of task-related attention demands on the corticofugal system seen in these experiments is specific to the motor network or spans the neural system as a whole. Future experiments measuring the auditory brainstem response should make use of a visual attention task that either does not require the participant to perform a motor behavior at all, or does not require a regular motor behavior in unison with the visual stimulus while presenting a background auditory stimulus to further determine whether processes engaged during a motor behavior specifically or in performing a non-auditory task in general control the descending auditory pathway. Jones and colleagues (2002) and Mathewson et al. (2010) have demonstrated behavioral differences in auditory and visual perception for rhythmically presented stimuli without the presence of a motor behavior, but neural evidence for these differences at the earliest levels of human perception are still lacking.

The parallels between the cortical EEG data and the brainstem response data presented in these experiments suggest that there may be a functional link between cortico-cortical interactions and the descending corticofugal projections to the auditory brainstem. However, the cortical EEG measurements presented in this dissertation were taken from the recordings of a single electrode placed at vertex (Cz). This is not sufficient to fully understand the interaction of cortical networks and oscillatory activity with the peripheral auditory pathway, as some cortical activity is specific to certain brain regions (e.g. mu rhythms in motor cortex, Chatrian et al., 1959; Howe & Sterman, 1972), some neural processes are signified by changes in power or synchrony in a certain frequency band across separate brain regions (e.g. theta and alpha band strength across frontal and association cortices during executive control tasks, von Stein & Sarnthein, 2000) and, most importantly for the current experiments, the effects of cognitive processes on cortical oscillation activity are often demonstrated across a large network of brain regions (i.e. attentional modulation of oscillatory entrainment occurs across a network of primary sensory, motor, association and frontal cortical regions, Besle et al., 2011). Future experiments should collect cortical EEG across the entire scalp while simultaneously measuring the auditory brainstem response in order to fully examine the relationship between fluctuations in cortical oscillation synchrony within and across multiple brain regions and the early transmission of acoustic input by the auditory brainstem. Since many accounts of cortical attention networks demonstrate an interaction between theta and alpha oscillations across the entire brain during executive control tasks, it is possible that the enhancement effect of the FFR in the auditory brainstem will correspond with changes in theta and alpha power across frontal, parietal and association cortices that indicate increased attentional and working memory resources to the auditory system. Such findings would suggest that the descending auditory pathway is subject to inputs from a broad attention network that monitors the allocation of resources across all modalities, rather than simply from cortical auditory or motor networks. Additionally, interactions between cortical motor and auditory networks could still demonstrate a suppression of primary auditory cortex by corollary discharge projections from the motor system while cross-modal attention networks still exhibit an increase in resources to the auditory system, indicating that the local suppression of the auditory cortex by the motor system may in fact be part of a larger network that enhances early sensory processing based on specific events or moments within a behavioral task, but then suppresses later processing of irrelevant sensory input in cortical sensory networks.

Even though the main focus of the experiments presented here was to examine the effects of cortical processes on early auditory perception, the brainstem response results taken with the cortical oscillation findings of these experiments suggest that these experiments may have implications for neural accounts of speech perception as well. Giraud and Poeppel (2012) propose a computational model of speech perception in which neural oscillations in the auditory cortex, particularly theta and gamma oscillations, entrain to the syllabic rhythm of speech to facilitate perception. Recent evidence from magnetoencephalography studies has supported this model by demonstrating that speech entrains the phase of delta and theta oscillations and the amplitude of gamma oscillations in the auditory cortex (Gross et al., 2013). The tuning of the brainstem response as well as cortical theta band oscillations to task events demonstrated in the experiments presented here indicate that the enhancement of the brainstem response at significant moments in time could play an important role in the perception of speech input, particularly in adverse listening conditions. Future experiments should measure cortical and brainstem EEG responses to continuous speech to begin to understand how these oscillatory processes within the auditory cortex in response to speech input may modulate transmission of acoustic signal in the auditory brainstem via the descending peripheral auditory pathway.

The experiments presented here asked two main questions about the interaction between cortical networks and the auditory brainstem via the corticofugal system. First, given evidence that the primary auditory cortex is suppressed by corollary discharge fibers from the motor system during movement, do these projections also suppress the processing of auditory input at the level of the brainstem? The results of all three experiments answer this question in the negative, as the introduction of a motor behavior or a motor response led to an enhancement of the frequency following response in the auditory brainstem. Second, the dissertation asked whether the motor suppression of the auditory response could be due to divided attention with a focus on the task relevant to the demands of the situation. Once again, all three experiments refuted this hypothesis by demonstrating either equal amounts of enhancement of the FFR during the motor behavior regardless of attentional task demands (Experiments 1 and 2) or increasing enhancement of the brainstem response to an auditory stimulus with increased task difficulty (Experiment 3). Overall, the results suggest that the descending auditory pathway may be modulated by a system that directs attention within all sensory networks to specific task events or moments to enhance early sensory perception. Furthermore, they demonstrate the need to include the corticofugal system in neural models of human sensory perception and indicate that future research in auditory and speech perception should examine the interaction between subcortical encoding and transmission of sensory input and the oscillatory activity of cortical networks.

References

Abbs, J. H., & Sussman, H. M. (1971). Neurophysiological feature detectors and speech perception: A discussion of theoretical implications. *Journal of Speech and Hearing Research, 14*, 23-36.

Adams, J. (1995). Sound stimulation induces Fos-related antigens in cells with common morphological properties through the auditory brainstem. *Journal of Comparative Neurology, 361*, 645-668.

Akimov, N. P., Marshak, D. W., Frishman, L. J., Glickman, R. D., & Yusupov, R. G. (2010). Histamine reduces flash sensitivity of ON ganglion cells in the primate retina. *Investigative Ophthalmology & Visual Science, 51*(7), 3825-3834.

Alsius, A., Navarra, J., Campbell, R., & Soto-Faraco, S. (2005). Audiovisual integration of speech falters under high attention demands. *Current Biology, 15*, 839-843.

Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *Journal of Cognitive Neuroscience, 15*. 600-609.

Besle, J., Schevon, C. A., Mehta, A. D., Lakatos, P., Goodman, R. R., McKhann, G. M., ... & Schroeder, C. E. (2011). Tuning of the human neocortex to the temporal dynamics of attended events. *The Journal of Neuroscience*, *31*(9), 3176-3185.

Blakemore, S. J., Wolpert, D., & Frith, C. (2000). Why can't you tickle yourself?. *Neuroreport, 11*, R11-16.

Boersma, Paul (2001). Praat, a system for doing phonetics by computer. *Glot International* 5:9/10, 341-345.

Brown, M. C. (2011). Anatomy of olivocochlear neurons. In D. K. Ryugo, & R. R. Popper (Eds.), *Auditory and Vestibular Efferents*, (pp. 17-37). New York: Springer Science & Business Media.

Chandrasekaran, B., Hornickel, J., Skoe, E., Nicol, T., & Kraus, N. (2009). Context-dependent encoding in the human auditory brainstem relates to hearing speech in noise: implications for developmental dyslexia. *Neuron*, *64*(3), 311-319.

Chandrasekaran, B., & Kraus, N. (2010). The scalp-recorded brainstem response to speech: Neural origins and plasticity. *Psychophysiology*, *47*(2), 236-246.

Chatrian, G. E., Petersen, M. C., & Lazarte, J. A. (1959). The blocking of the rolandic wicket rhythm and some central changes related to movement. *Electroencephalography and Clinical Neuophysiology, 11*, 497-510.

Conway, B. A., Halliday, D. M., Farmer, S. F., Shahani, U., Maas, P., Weir, A. I., & Rosenberg, J. R. (1995). Synchronization between motor cortex and spinal motoneuronal pool during the performance of a maintained motor task in man. *The Journal of Physiology*, *489*(3), 917-924.

Cooper, N. P., & Guinan Jr, J. J. (2006). Medial olivocochlear efferent effects on basilar membrane responses to sound. In A. L. Nuttall, T. Ren, P. G. Gillespie, K. Grosh, & E. de Boer (eds.). *Auditory Mechanisms: Processes and Models,* (pp. 86-92). Singapore: World Scientific.

Davis, M. H. & Johnsrude, I. S. (2007). Hearing speech sounds: Top-down influences on the interface between audition and speech perception. *Hearing Research, 229*, 132-147.

de Boer, J., & Thornton, A. R. (2008). Neural correlates of percpetual learning in the auditory brainstem: Efferent activity predicts and reflects improvement at a specch-in-noise discrimination task. *Journal of Neuroscience, 28*, 4929-4937.

Deiber, M. P., Missonnier, P., Bertrand, O., Gold, G., Fazio-Costa, L., Ibanez, V., & Giannakopoulos, P. (2007). Distinction between perceptual and attentional processing in working memory tasks: a study of phase-locked and induced oscillatory brain dynamics. *Journal of Cognitive Neuroscience*, *19*(1), 158-172.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience, 18*, 193-222.

Diehl, R. L., Kluender, K. R., Walsh, M. A., & Parker, E. M. (1991). Auditory enhancement in speech perception and phonology. In R.R. Hoffman and D.S. Palermo (Eds.), *Cognition and the Symbolic Processes: Applied and Ecological Perspectives* (pp. 59-76). New York, NY: Psychology Press.

Dolan, D. F., & Nuttall, A. L. (1988). Masked cochlear whole-nerve response intensity functions altered by electrical stimulation of the crossed olivocochlear bundle. *Journal of the Acoustical Society of America, 83*, 1081-1086.

Eliades, S. J., & Wong, X. (2008). Neural substrates of vocalization feedback monitoring in primate auditory cortex. *Nature, 453*, 1102-1107.

Engel, A. K., & Fries, P. (2010). Beta-band oscillations—signaling the status quo?. *Current Opinion in Neurobiology*, *20*(2), 156-165.

Engel, A. K., Fries, P., Singer, W. (2001). Dynamic predictions: Oscillations and synchrony in top-down processing. *Nature Reviews, 2*, 704-716.

Eriksen, C. W., & Eriksen, B. A. (1974). Effects of noise letters upon the identification letter in a non-search task. *Perception and Psychophysics, 16*, 143-149.

Eriksen, C. W., & Schultz, D. W. (1979). Information processing in visual search: a continuous flow conception and experimental results. *Perception and Psychophysics, 25*, 249-263.

Foxe, J. J., & Snyder, A. C. (2011). The role of alpha-band brain oscillations as a sensory suppression mechanism during selective attention. *Frontiers in Psychology*, *2*, 1-13.

Friederici, A. D. (2012). The cortical language circuit: From auditory perception to sentence comprehension. *Trends in Cognitive Sciences, 16*, 262-268.

Galbraith, G. C. & Arroyo, C. (1993). Selective attention and brainstem frequency-following responses. *Biological Psychology, 37*, 3-22.

Galbraith, G. C., Bhuta, S. M., Choate, A. K., Kitahara, J.M., & Mullen Jr., T. A. (1998). Brain stem frequency-following response to dichotic vowels during attention. *NeuroReport, 9*, 1889-1893.

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.

Galbraith, G. C., Olfman, D. M., & Huffman, T. M. (2003). Selective attention affects human brain stem frequency-following response. *NeuroReport, 14*(5), 735-738.

Galbraith, G. C., Threadgill, M. R., Hemsley, J., Salour, K., Songdej, N., Ton, J., & Cheung, L. (2000). Putative measure of peripheral and brainstem frequency-following in humans. *Neuroscience Letters*, *292*(2), 123-127.

Garinis, A. C., Glattke, T., & Cone, B. K. (2011). The MOC reflex during active listening to speech. *Journal of Speech, Language and Hearing Research, 54*, 1464-1476.

Gastinger, M. J., Tian, N., Horvath, T., & Marshak, D. W. (2006). Retinopetal axons in mammals: Emphasis on histamine and serotonin. *Current Eye Research, 31*(7-8), 655-667.

Giard, M., Collet, L., Bouchet, P., & Pernier, J. (1993). Auditory selective attention in the human cochlea. *Brain Research, 633*, 353-356.

Giraud, A. L., & Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nature neuroscience*, *15*(4), 511-517.

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. *PLoS Biol*, *11*(12), e1001752.

Guinan Jr, J. J. (2006). Olivocochlear efferents: Anatomy, physiology, function, and the measurement of efferent effects in humans. *Ear & Hearing, 27*(6), 589-607.

Hairston, W. D., Letowski, T. R., & McDowell, K. (2013). Task-related suppression of the brainstem frequency following response. *PLOS One, 8*(2), e55215.

Hashikawa, T. (1983). The inferior colliculopontine neurons of the cat in relation to other collicular descending neurons. *Journal of Computational Neurology, 219,* 241-249.

Heald, S. L. M., & Nusbaum, H. C. (2014). Speech perception as an active cognitive process. *Frontiers in Systems Neuroscience, 8*, 1-15.

Hickok, G., Houde, J., & Rong, F. (2011). Sensorimotor integration in speech processing: Computational basis and neural organization. *Neuron, 69*(3), 407-422.

Hickok, G. & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience, 8*, 393-402.

Holt, L. L., & Lotto, A. J. (2008). Speech perception within an auditory cognitive science framework. *Current Directions in Psychological Science, 17*(1), 42-46.

Hood, L. (1998). *Clinical Applications of the Auditory Brainstem Response*. San Diego, CA: Singular Publishing Group.

Hoormann, J., Falkenstein, M., & Hohnsbein, J. (1994). Effect of selective attention on the latency of human frequency-following potentials. *NeuroReport, 5*, 1609-1612.

Hoormann, J., Falkenstein, M., & Hohnsbein, J. (2004). Effects of spatial attention on the brain stem frequency-following potential. *NeuroReport, 15*(10), 1539-1542.

Houde, J.F., Nagarajan, S. S., Sekihara, K., & Merzenich, M. M. (2002). Modulation of the auditory cortex during speech: An MEG study. *Journal of Cognitive Neuroscience, 14*(8), 1125-1138.

Howe, R. C., & Sterman, M. B. (1972). Cortical-subcortical EEG correlates of suppressed motor behavior during sleep and waking in the cat. *Electroencephalography and Clinical Neurophysiology, 32*, 681-695.

Huffman, R. F., & Henson Jr., O. W. (1990). The descending auditory pathway and acousticomotor systems: Connections with the inferior colliculus. *Brain Research Reviews, 15*, 295-323.

Johnson, J. A., & Zatorre, R. J. (2006). Neural substrates for dividing and focusing attention between simultaneous auditory and visual events. *Neuroimage*, *31*(4), 1673-1681.

Jones, M. R., Moynihan, H., MacKenzie, N., & Puente, J. (2002). Temporal aspects of stimulus-driven attending in dynamic arrays. *Psychological Science*, *13*(4), 313-319.

Kawase, T., Delgutte, B., & Liberman, M. C. (1993). Anti-masking effects of the olivocochlear reflex, II: Enhancement of auditory-nerve response to masked tones. *Journal of Neurophysiology, 70*, 2533-2549.

Kay, L.M., & Freeman, W. J. (1998). Bidirectional processing in the olfactory-limbic axis during olfactory behavior. *Behavioral Neuroscience, 112*(3), 541-553.

Klatt, D. H. (1980). Software for a cascade/parallel formant synthesizer. *The Journal of the Acoustical Society of America*, *67*(3), 971-995.

Kleinfeld, D., Ahissar, E., & Diamond, M. E. (2006). Active sensation: insights from the rodent vibrissa sensorimotor system. *Current Opinion in Neurobiology*, *16*(4), 435-444.

Krishnan, A. (2002). Human frequency-following responses: representation of steady-state synthetic vowels. *Hearing Research, 166*, 192-201.

Krishnan, A. (2007). Frequency-following response. In R. F. Burkard, M. Don & J. J. Eggermont (Eds.), *Auditory Evoked Potentials: Basic Principles and Clinical Application*, (pp. 313-331), Baltimore, MD: Lippincott Williams & Wilkins.

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*, 161-168.

Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review*, *106*(1), 119-159.

Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology: Human Perception and Performance, 21*(3), 451-468.

Lavie, N. (1997). Feature integration and selective attention: Response competition from unattended distractor features. *Perception & Psychophysics, 59,* 543-556.

Lee, J. H., Whittington, M. A., & Kopell, N. J. (2013). Top-down beta rhythms support selective attention via interlaminar interaction: A model. *PLoS Computational Biology*, *9*(8), e1003164.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74*(6), 431-461.

Lundqvist, M., Rose, J., Herman, P., Brincat, S. L., Buschman, T. J., & Miller, E. K. (2016). Gamma and beta bursts underlie working memory. *Neuron, 90*, 1-13.

Maison, S., Micheyl, C., & Collet, L. (2001). Influence of focused auditory attention on cochlear activity in humans. *Psychophysiology, 38*, 35-40.

Martikainen, M. K., Kaneko, K., & Hari, R. (2005). Suppressed responses to self-triggered sounds in the human auditor cortex. *Cerebral Cortex, 15*(3), 299-302.

Mathewson, K. E., Fabiani, M., Gratton, G., Beck, D. M., & Lleras, A. (2010). Rescuing stimuli from invisibility: Inducing a momentary release from visual masking with pre-target entrainment. *Cognition*, *115*(1), 186-191.

MATLAB version 8.4.0. Natick, Massachusetts: The MathWorks Inc., 2014.

Milner, A.D., & Goodale, M. A. (1995). *The visual brain in action.* Oxford: Oxford University Press.

Moore, J. K. (1987). The human auditory brain stem: A comparative view. *Hearing Research, 29*(1), 1-32.

Morillon, B., Hackett, T. A., Kajikawa, Y., & Schroeder, C. E. (2015). Predictive motor control of sensory dynamics in auditory active sensing. *Current Opinion in Neurobiology, 31*, 230-238.

Musacchia, G., Sama, M., Skoe, E., & Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *PNAS, 104*(40), 15894-15898.

Newman, S. D., Keller, T. A., & Just, M. A. (2007). Volitional control of attention and brain activation in dual task performance. *Human Brain Mapping*, *28*(2), 109-117.

Nusbaum, H. C., & Magnuson, J. S. (1997). Talker normalization: Phonetic constancy as a cognitive process. In K. Johnson and J. W. Mullennix (Eds.), *Talker Variability in Speech Processing*, pp. 109-132. San Diego: Academic Press.

Nusbaum, H. C., & Schwab, E. C. (1986). The role of attention and active processing in speech perception. In E. C. Schwab and H. C. Nusbaum (Eds.) *Pattern Recognition by Humans and Machines: Speech Perception* (pp. 113-157). San Diego, CA: Academic Press.

Papanicolaou, A. C., Raz, N., Loring, D. W., & Eisenberg, H. M. (1986). Brain stem evoked response suppression during speech production. *Brain and Language, 27*, 50-55.

Pfurtscheller, G., Stancak, A., & Neuper, C. (1996). Post-movement beta synchronization. A correlate of an idling motor area? *Electroencephalography and Clinical Neurophysiology*, *98*(4), 281-293.

Picton, T. W., Stapells, D. R., & Campbell, K. B. (1981). Auditory evoked potentials from the human cochlea and brainstem. *The Journal of Otolaryncology, 9*, 1-41.

Raghavachari, S., Kahana, M. J., Rizzuto, D. S., Caplan, J. B., Kirschen, M. P., Bourgeois, B., ... & Lisman, J. E. (2001). Gating of human theta oscillations by a working memory task. *TheJjournal of Neuroscience*, *21*(9), 3175-3183.

Rasmussen, G. L. (1946). The olivary peduncle and other fiber projections of the superior olivary complex. *Journal of Comparative Neurology, 99*, 61-74.

Rauschecker, J. P. & Scott, S. K. (2009). Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nature Neuroscience, 12*, 718-724.

Recanzone, G. H., & Sutter, M. L. (2008). The biological basis of audition. *Annual Review of Psychology, 59*, 119-142.

Resch, H., Zawinka, C., Lung, S., Weigert, G., Schmetterer, L., & Garhofer, G. (2005). Effect of histamine and cimetidine on retinal and choroidal blood flow in humans. *American Journal of Physiology: Regulatory, Integrative and Comparative Physiology, 289*, R1387-R1391.

Roux, F., & Uhlhaas, P. J. (2014). Working memory and neural oscillations: alpha-gamma versus theta-gamma codes for distinct WM information?. *Trends in Cognitive Sciences, 18*(1), 16-25.

Sanes, J. N., & Donoghue, J. P. (1993). Oscillations in local field potentials of the primate motor cortex during voluntary movement. *Proceedings of the National Academy of Sciences*, *90*(10), 4470-4474.

Sauseng, P., Klimesch, W., Schabus, M., & Doppelmayr, M. (2005). Fronto-parietal EEG coherence in theta and upper alpha reflect central executive functions of working memory. *International Journal of Psychophysiology, 57*, 97-103.

Schneider, D. M., Nelson, A., & Mooney, R. (2014). A synaptic and circuit basis for corollary discharge in the auditory cortex. *Nature, 513*, 189-194.

Schwartz, I. R. (1992). The superior olivary complex and lateral lemniscal nuclei. In D. B. Webster and R. R. Fay (Eds.), *The Mammalian Auditory Pathway: Neuroanatomy*, pp. 117-167. New York, NY: Springer-Verlag.

Sherman, S.M. (2006). What is the function of the thalamus? In J. L. van Hemmen and T. J. Sejnowski (Eds.), *23 Problems in Systems Neuroscience*, pp. 65-82. Oxford University Press.

Sherman, S.M. (2009). Thalamocortical relations. In G.G. Bernston and J. T. Cacioppo (Eds.), *Handbook of Neuroscience for the Behavioral Sciences, Volume 1*, pp 201-223. Hoboken, NJ: Wiley & Sons.

Skipper, J. I. (2014). Echoes of the spoken past: How auditory cortex hears context during speech perception. *Philosophical Transactions of The Royal Society Biological Sciences, 369*(1651), 20130297.

Skipper, J. I., van Wassenhove, V., Nusbaum, H. C., & Small, S. L. (2007). Hearing lips and seeing voices: How cortical areas supporting speech production mediate audiovisual speech perception. *Cerebral Cortex*, *17*(10), 2387-2399.

Skoe, E., & Kraus, N. (2010). Hearing it again and again: on-line subcortical plasticity in humans. *PLoS One*, *5*(10), e13645.

Skoe, E., Krizman, J., Spitzer, E., & Kraus, N. (2013). The auditory brainstem is a barometer of rapid auditory learning. *Neuroscience*, *243*, 104-114.

Sperry, R. W. (1950). Neural basis of the spontaneous optokinetic response produced by visual inversion. *Journal of comparative and physiological psychology*, *43*(6), 482.

Spinelli, D.N., & Pribram, K. H. (1966). Changes in visual recovery functions produced by temporal lobe stimulation in monkeys. *Electroencephalography and Clinical Neurophysiology, 20*(1), 44-49.

Steriade, M., McCormick, D. A., & Sejnowski, T. J. (1993). Thalamocortical oscillations in the sleeping and aroused brain. *Science*, *262*(5134), 679-685.

Stevens, K. N. (1960). Toward a model for speech recognition. *The Journal of the Acoustical Society of America, 32*(1), 47-55.

Stevens, K.N., & Halle, M. (1967). Remarks on analysis by synthesis and distinctive features. In W. Walthen-Dunn (Ed.), *Models for the perception of speech and visual form* (pp. 88-102). Cambridge, MA: MIT Press.

Suga, N., & Shimozawa, T. (1974). Site of neural attenuation of responses to self-vocalized sounds in echolocating bats. *Science, 183*(4130), 1211-1213.

Suga, N., Xiao, Z., Ma, X., & Ji, W. (2002). Plasticity and corticofugal modulation for hearing in adult animals. *Neuron, 36*, 9-18.

Swann, N., Tandon, N., Canolty, R., Ellmore, T. M., McEvoy, L. K., Dreyer, S., ... & Aron, A. R. (2009). Intracranial EEG reveals a time-and frequency-specific role for the right inferior frontal gyrus and primary motor cortex in stopping initiated responses. *The Journal of Neuroscience*, *29*(40), 12675-12685.

Tierney, A., & Kraus, N. (2013). The ability to move to a beat is linked to the consistency of neural responses to sound. *The Journal of Neuroscience, 33*(38), 14981-14988.

Thompson, A. M., & Thompson, G. C. (1991). Posteroventral cochlear nucleus projections to olivocochlear neurons. *Journal of Comparative Neurology, 303*, 267-285.

Ungerleider, L.G., & M. Mishkin. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, and R. J. W. Mansfield (Eds.). *Analysis of Visual Behavior* (pp. 549-586). Cambridge, Mass.: MIT Press.

Von Holst, E. (1954). Relations between the central nervous system and the peripheral organs. *The British Journal of Animal Behaviour*, *2*(3), 89-94.

von Stein, A., & Sarnthein, J. (2000). Different frequencies for different scales of cortical integration: from local gamma to long range alpha/theta synchronization. *International Journal of Psychophysiology, 38*, 301-313.

Winslow, R. L., & Sachs, M. B. (1988). Single-tone intensity discrimination based on auditory-nerve rate responses in backgrounds of quiet, noise and with stimulation of crossed olivo-cochlear bundle. *Hearing Research, 35*, 165-190.

Wolfe, J. M., Kluender, K. R., Levi, D. M., Bartoshuk, L. M., Herz, R. S., Klatzky, R. L., Lederman, S. J., & Merfeld, D. M. (2012). *Sensation and Perception* (3rd ed.). Sunderland, MA: Sinauer Associates, Inc.

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.

Appendix A: Musical Experience Questionnaire

1. Do you play (or have you played) any musical instruments? YES NO

   1b. At what age (approximately) did you first begin musical instruction? _________

   1c. Do you still actively play a musical instrument? YES NO

2. If you circled YES in Question 1, please list all the musical instruments you have played, including the number of years you have played them:

Instrument____________________________________ Number of Years_________

Instrument____________________________________ Number of Years_________

Instrument____________________________________ Number of Years_________

Instrument____________________________________ Number of Years_________

Instrument____________________________________ Number of Years_________

Voice____________________________________ Number of Years_________

3. Do you know how to read music? YES NO

4. Have you ever played in any ensembles or sung in any choirs? YES NO

5. If you answered YES to Question 4, please list the ensembles and/or choirs with which you played/sung.

Ensemble/Choir________________________________ Number of Years_________

Ensemble/Choir________________________________ Number of Years_________

6. Have you taken any music courses at the high school or collegiate level? YES NO

7. If you answered YES to Question 6, please list the approximate number of music courses you have completed

   ____

8. Do you have perfect pitch (the ability to name any note without a reference)?

   YES NO

9. Approximately how many hours do you spend making music per week? (circle one):

   Less than one hour 1-2 hours 3-4 hours 5-6 hours More than 6 hours

10. Approximately how many hours do you spend listening to music per day (circle one):

    Less than one hour 1-2 hours 3-4 hours 5-6 hours More than 6 hours

11. Approximately what percentage of music you listen to falls under the following genres?

    Pop/Rock ________

    Dance/Techno ________

    Jazz ________

    R&B ________

    Hip Hop ________

    Classical ________

    Folk ________

    Non-Western/Ethnic ________

    Other (please specify) ________

12. Approximately how many concerts do you attend per year? ______