ABSTRACT

Title of Document: HEARING VS. LISTENING:
ATTENTION CHANGES THE NEURAL
REPRESENTATIONS
OF AUDITORY PERCEPTS

Juanjuan Xiang, Ph.D., 2008

Directed By: Professor Jonathan Z. Simon, Department of
Electrical and Computer Engineering

Making sense of acoustic environments is a challenging task. At any moment, the signals from distinct auditory sources arrive in the ear simultaneously, forming an acoustic mixture. The brain must represent distinct auditory objects in this complex scene and prioritize processing of relevant stimuli while maintaining the capability to react quickly to unexpected events. The present studies explore neural representations of temporal modulations and the effects of attention on these representations. Temporal modulation plays a significant role in speech perception and auditory scene analysis. To uncover how temporal modulations are processed and represented is potentially of great importance for our general understanding of the auditory system.

Neural representations of compound modulations were investigated by magnetoencephalography (MEG). Interaction components are generated by near rather

than distant modulation rhythms, suggesting band-limited modulation filter banks operating in the central stage of the auditory system. Furthermore, the slowest detectable neural oscillation in the auditory cortex corresponds to the perceived oscillation of the auditory percept.

Interactions between stimulus-evoked and goal-related neural responses were investigated in simultaneous behavioral-neurophysiological studies, in which we manipulate subjects' attention to different components of an auditory scene. Our experimental results reveal that attention to the target correlates with a sustained increase in the neural target representation, beyond well-known transient effects. The enhancement of power and phase coherence presumably reflects increased local and global synchronizations in the brain. Furthermore, the target's perceptual detectability improves over time (several seconds), correlating strongly with the target representation's neural buildup. The change in cortical representations can be reversed in a short time-scale (several minutes) by various behavioral goals.

These aforementioned results demonstrate that the neural representation of the percept is encoded using the feature-driven mechanisms of sensory cortex, but shaped in a sustained manner via attention-driven projections from higher-level areas. This adaptive neural representations occur on multiple time scales (seconds vs. minutes) and multiple spatial scales (local vs. global synchronization). Such multiple resolutions of adaptation may underlie general mechanisms of scene organization in any sensory modality and may contribute to our highly adaptive behaviors.

HEARING VS. LISTENING:
ATTENTION CHANGES THE NEURAL REPRESENTATIONS OF AUDITORY PERCEPTS

by

Juanjuan Xiang

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2008

Advisory Committee:

Professor Jonathan Z. Simon, Chair
Professor Shihab A. Shamma
Professor William Levine
Professor Monita Chatterjee
Professor Cynthia F. Moss

DEDICATION

To my parents Zuoxian Xiang and Aiying Liu for their unconditional love

ACNOWLEDGEMENTS

I would like to express my gratitude to all those who gave me the possibility to complete this thesis.

First of all, I would like thank my advisor, Professor Jonathan Z. Simon, for providing me with the opportunity to pursue the doctorate degree and for always being accessible through the research life. His attitudes of detail-orientation and persistence are worthy to learn and follow. Special thoughts go also to Professor Mounya Elhilali and Professor Shihab Shamma. The main part of the thesis has been a delightful and fruitful collaboration with them. To Professor David Poeppel, who was always willing to advise and help.

I would also like to thank the past and present members of the Computational Sensorimotor Systems Lab, the Neural Systems Laboratory and the Cognitive Neuroscience of Language lab: Nayef Ahmar, Pingbo Yin, Nima Mesgarani, Ling Ma, Serin Atiani, Jeff Walker, Huan Luo, Yadong Wang and everybody else. Thank you for helping me in so many ways.

I was fortunate to have around a group of cool friends that were always there to share my pleasure and ease my anxiety. Thanks Xing Tian, Haiyan He, Feng Rong, Jin Liang, Kaiwen He, Zhen Shi, Xiaosong Hu and Youjun Wang, for sharing so many wonderful time together.

My deepest gratitude goes to my family for their unconditional love and support throughout my life. They gave all they had to provide the best possible environment for me to grow up, to get education and to become a man of independence and integrity.

Although I did not have much time to accompany them in the past five years, my love is always with them.

College Park, Maryland,

April, 2008

Juanjuan Xiang

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1: Introduction

Making sense of acoustic environments is a challenging task. At any moment, the signals from distinct auditory sources arrive in the ear simultaneously, forming an acoustic mixture. The brain must recover distinct auditory objects from this complex scene and prioritize processing of relevant stimuli while maintaining the capability to react quickly to unexpected events. This complex process is referred to as auditory scene analysis (ASA). How does the brain accomplish this feat? What are the critical features extracted and encoded by the brain to organize the percept? How do cognitive functions, such as behavioral goals or attention, affect the representations of these features?

Over the past few decades, a large body of research in neuroscience has been aimed at answering these questions. Emerging evidence suggests a hierarchical organization of sound feature processing. Ascending the neural axis of the auditory system, neurons extract increasingly complex sound features, ranging from frequency, to location, to temporal modulation. High-level cognitive functions, such as attention, emotion, and life experience, affect perception by shaping the neural representations of features.

The present studies explore neural representations of temporal modulation and the modulatory effect of attention on these representations. Temporal modulation plays a significant role in speech perception and ASA, for example in stream segregation

(Bregman, 1990) and signal detection in the presence of maskers (Hall & Harvey, 1984). Thus, to uncover how temporal modulations are processed and represented is potentially of great importance for our general understanding of the auditory system.

This dissertation is organized in five chapters. In the first chapter we review the present knowledge of cortical representations of features and the modulatory effects of cognitive functions. Chapter 2 examines representations of multiple temporal modulations, an essential feature of vocalization and speech but one that has been inadequately investigated. Chapter 3 and chapter 4 explore the effects of attention on cortical representations of temporal modulation using two different paradigms, known as 'informational masking' and 'stream segregation', respectively. In the former paradigm, the target stream is embedded in maskers, while in the latter case two streams are presented concurrently, creating a sound rivalry situation. A summary of the main findings of this work and future prospects are presented in chapter 5.

The remainder of this chapter is organized as follows: The next section describes psychoacoustic studies which demonstrate the important role of temporal modulation in ASA and speech recognition. The second and third sections discuss related work on representations of features and the modulatory effect of cognitive functions on these representations, both from the perspective of single neurons and at the neuron population level. Section 1.4 discusses related fMRI studies, which provide complementary localization information about feature representations. The last section summaries the related work at various neural levels.

1.1 Psychoacoustic research

Natural sounds, including animal vocalizations and human speech, are time-varying signals. Information contained in the dynamic temporal structure is crucial for content identification and communication. There are two primary temporal scales: fast pressure variations define the fine structure of the sound, while the overall slow amplitude changes define the envelope (Viemeister & Plack, 1993). Our studies focus on the slow dynamic change. Human auditory system has a striking ability to detect temporal modulations. This is characterized by the temporal modulation transfer function (tMTF). The tMTF is determined by measuring threshold for detection of modulation of sinusoidally modulated noise at each modulation rate. Modulated white noise has a constant long term spectrum. So the only available cue for detection comes from temporal information (Bacon & Viemeister, 1985). The low pass shape of tMTF curve shows that sensitivity for modulation decreases progressively with increasing modulation rate. The dynamics of the envelope play an important role for auditory scene analysis and speech perception, which is discussed in detail in the section 1.1.1 and section 1.1.2. The neural correlates of the slow temporal modulation are discussed in the reminder of the chapter.

1.1.1 Auditory scene analysis

ASA refers to a segregation process by which relevant components are identified and attended to in the presence of interfering sounds in a complex auditory scene. Following one person’s conversation or a melody during a noisy cocktail party are two everyday examples of ASA. The physical cues that are critical for performing the ASA have been well established in the past few decades. There are two broad classes of segregation, concurrent segregation and sequential segregation. Here we consider only the latter.

Sequential segregation, or streaming, concerns the perceptual organization of sounds over time. Widely used stimuli are sequences of alternating pure tones (A and B tones), which can be perceived as either one or two streams, depending on several physical stimulus parameters (Figure 1.1). Small frequency and large temporal differences between A and B (ΔF, ΔT) and short duration of the sequence (T) promote the perceptual grouping of the tone sequence into one coherent stream. Conversely, for sounds with large ΔF, small ΔT, and long duration, the percept is split into two streams, one formed by the A tones, the other by the B tones (Bregman & Campbell, 1971; van Noorden, 1977). Tone sequences with variables between boundaries elicit an ambiguous and bi-stable percept (Pressnitzer & Hup, 2005). A channeling theory, which postulates that sounds exciting the same auditory channels are perceived as a single stream while sounds exciting different peripheral auditory channels are assigned to different perceptual streams, was proposed to account for these observations (Hartmann & Johnson, 1991).

Following these early demonstrations of auditory streaming, several studies have shown that streaming can be elicited by differences in other parameters, such as fundamental frequency (F0) (Vliegen & Oxenham, 1999) and modulation rate (Grimault, Micheyl, Carlyon, & Collet, 2002). In these studies, the stimuli were carefully designed to occupy the same frequency region, so that the same peripheral auditory channels are excited. The preservation of streaming suggests that central auditory nuclei play a functional role in ASA.

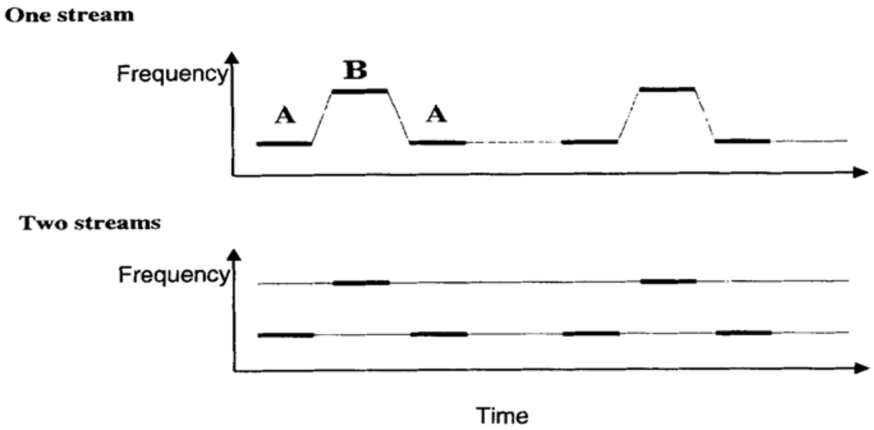

Figure 1.1: Schematic representation of the alternating tones paradigm. The dashed lines in the top and bottom panels of the figure demonstrate the perceptual organizations corresponding to one and two streams, respectively (van Noorden, 1975).

Recently, another paradigm, informational masking, was proposed to study the segregation occurring at a higher level in the auditory system than the periphery (Kidd Jr, Mason, Deliwala, Woods, & Colburn, 1994). A fixed-frequency target tone (target) is presented simultaneously with other spectral components (maskers). Detection of the target note is hindered by the maskers, which constitute a form of informational masking (Neff & Green, 1987). However, if the masker frequencies vary from one burst to the next within the course of the stimulus sequence (Figure 1.2), detection thresholds are substantially reduced, a signature of streaming.

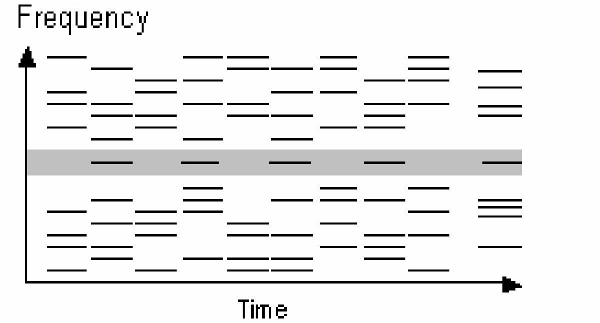

Figure 1.2: Schematic representation of informational masking paradigm. A regular tone sequence is embedded in a complex background. The gray rectangle depicts the protection zone, the minimal spectral difference between the target tone and the nearest markers.

1.1.2 Speech perception

One fundamental issue in speech perception involves identifying the perceptual unit into which continuous speech is broken. The principal difficulty in attempting to answer this question lies in co-articulation, which refers to the phenomenon that the production of each sound is affected by its context. Due to co-articulation, individual sounds cannot be isolated; hence, parsing out the perceptual unit directly is difficult. Accumulating psychoacoustic evidence suggests that two perceptual units exist (Poeppel, 2003). One is the syllabic unit (Greenberg, Carvey, Hitchcock, & Chang, 2003), corresponding to slow modulation (~ 4 Hz); another is phonetic segmentation (Liberman & Mattingly, 1985), corresponding to fast modulation (~ 25 Hz). Speech perception operates on both modulation rates simultaneously. Our first experiment, detailed in chapter two, investigates the auditory cortical representations for concurrent slow modulation and fast modulation.

Another essential feature of speech perception is its robustness. Recognition of speech is highly robust against many kinds of distortion, such as room reverberation, noise influence, and gap inserting. A study by Shannon et al. (1995) sought to identify cues that convey phoneme identity reliably under noise conditions. In this study, a speech signal was filtered into 4 frequency bands, and then the envelope components were extracted using the Hilbert transform. The temporal envelopes of the speech signals were used to modulate noise of the same bandwidths. High speech recognition performance was obtained with only three bands. These results demonstrate that presentation of a dynamic temporal pattern in only a few broad spectral regions is sufficient for the recognition of speech; the most critical information about speech content is preserved in the slow temporal envelope (< 20 Hz) (Drullman, Festen, & Plomp, 1994; Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995).

1.2 Traditional neurophysiology: single neuron recording

The hierarchical organization of feature processing assumes that the auditory system extracts progressively abstract features of a busy scene, ranging from frequency (in the inner ear) to complex spectral and temporal features (in the auditory cortex). To determine whether a particular feature is encoded in the auditory system, the activity of single neurons is recorded in response to stimuli with parametrically varied quantities on a particular feature dimension while maintaining other features constant. Research with single-neuron recording using this paradigm has identified several critical features to which neurons are tuned, such as frequency (Talavage et al., 2004), pitch (Bendor & Wang, 2005), intensity (Bilecen, Seifritz, Scheffler, Henning, & Schulte, 2002), modulation rate (Heil & Irvine, 1998), and speech-specific features (Ahveninen et al., 2006). An individual neuron is receptive to auditory stimuli with particular physical features, i.e., each neuron is tuned to a narrow range of values of that feature. Thus plotting the best feature value of all neurons forms a feature map. Research on neural responses to modulation is detailed in section 1.2.1.

Hierarchically organized feature maps alone cannot explain all cognitive processes, such as perceptual learning. Much evidence has suggested an important role of short-term plasticity, i.e., feature maps are not stationary. Instead, they are dynamically modulated by cognitive functions, such as stimulus context, behavior context, and attention. The plasticity of feature maps will be discussed in section 1.2.2.

Single-neuron recordings provide direct measures of microscopic activities in the brain and lay a solid foundation for further understanding of the complex auditory system. However, there are two main limitations of this technique. First, it is an invasive technique and cannot be applied to healthy human beings due to ethical restrictions. Thus it is hard, if not impossible, to implement this technique to investigate high-level cognitive functions and language processing. Second, it is difficult to investigate the activities at the systems level, where different encoding mechanisms might be employed among neural ensembles. Encoding mechanisms at the systems level reflect macroscopic activities, and thus might represent better predictions of behavior. Section 1.3 and section 1.4 discuss related work using brain imaging techniques, which are non-invasive and measure activities of neural populations; therefore, these techniques overcome the two aforementioned limitations.

1.2.1 Representations of temporal modulation

Two measurements are commonly used in neurophysiology studies to characterize neural responses. One is the average discharge rate (i.e., the number of spikes evoked over several modulation cycles), known as ‘rate coding’. Another is the vector strength, a measure of synchronization of the timing of action potentials to the envelope waveform, known as ‘temporal coding’. Widely used modulated stimuli include click trains, square modulated tone bursts, sinusoidally amplitude modulated (SAM) noise, and beat sounds (Joris, Schreiner, & Rees, 2004).

Temporal modulation information is represented by temporal coding in the auditory periphery. Take SAM tones as an example (Figure 1.3 A), where a high frequency tone (carrier) is modulated by a low frequency signal (envelope). The mathematical form is in Equation 1, where m denotes modulation depth, f_m modulation frequency, f_c carrier frequency and t time. Figure 1.3 B shows the spectrum. Figure 1.3 C shows a post-stimulus time histogram (PSTH)

$$s(t) = [1 + m\sin(2\pi f_m t)]\sin(2\pi f_c t) \qquad \text{(Eq. 1)}$$

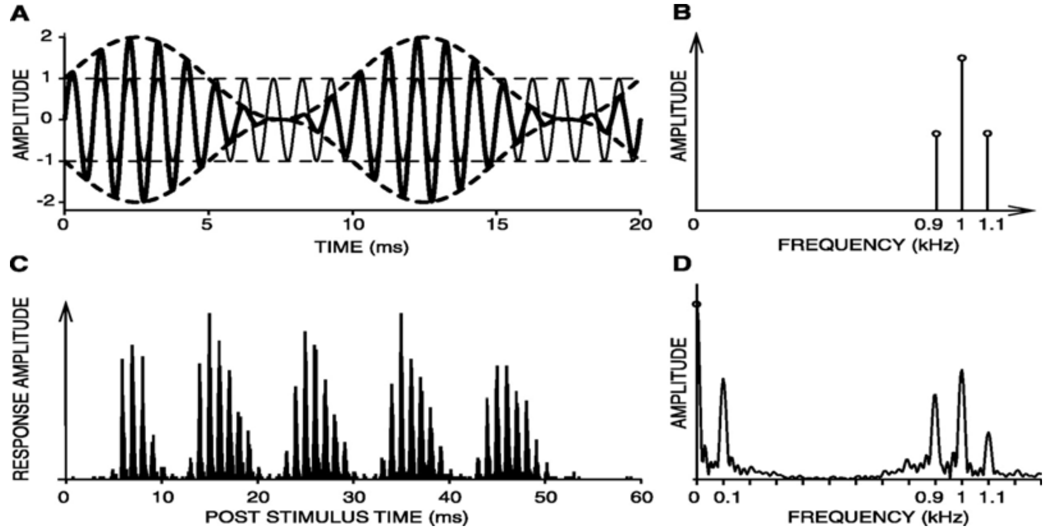

Figure 1.3: A: Superimposed waveforms of a 1 KHz tone (thin line) and the same tone sinusoidally amplitude modulated (AM) (thick line) at 100% with a modulation frequency of 100 Hz. B: Spectrum of the AM tone in A. C: Post-stimulus time (PST) histogram of a nerve

of an auditory nerve (AN) to the stimulus in figure 1.3 A (Joris, Schreiner, & Rees, 2004). The response phase locks to both the fine structure and the envelope of that stimulus in figure 1.3A. Phase-locking is also demonstrated in the spectrum of the PSTH; phase-locking to the fine structure is indicated by the f_c response, while phase-locking to the envelope is indicated by the f_m component (figure 1.3 D). The component at f_m is represented in the response despite its absence in the stimuli (Figure 1.3 D, Figure 1.3 A).

The representations of temporal modulation at more central stages of the auditory system have been investigated in the monkey (Liang, Lu, & Wang, 2002; Malone, Scott, & Semple, 2007; Wang, Lu, & Liang, 2003) the gerbil (Schulze & Langner, 1997); the cat (Eggermont, 1999); and the fish (Benda, Longtin, & Maler, 2006; Bodnar & Bass, 1997). Three basic principles are emerging from this body of work. First, ascending the auditory axis, the highest valid modulation rate decreases, ending with upper cut-off frequencies around 80 Hz in auditory cortex, the uppermost stage of the auditory system (Bieser & Muller-Preuss, 1996; Eggermont, 1998; Schreiner & Urbas, 1988). Second, in auditory cortex, slow modulation information (< 20 Hz) is encoded by synchronized neuron populations using temporal coding, while relatively fast modulation (20 Hz < f <50 Hz) is encoded by non-synchronized neuron populations using rate coding (Liang, Lu, & Wang, 2002; Wang, Lu, & Liang, 2003). Last, a high percentage of temporal-coding neurons show band pass characteristics, i.e., neurons tune to distinct modulation rates. Thus, the neuron population forms a periodotopy in addition to the classical tonotopy (Gaese & Ostwald, 1995; Lu, Liang, & Wang, 2001).

Recently, two new trends have appeared in single-neuron recording research. One involves using more sophisticated stimuli to probe the auditory cortex. Using ripple stimuli (noise or tone carriers with spectrotemporally modulated envelopes), Elhilali et al. (2004) showed that cortical neurons can track both envelope and fine structure. Furthermore, responses to slowly modulated envelopes regulated the expression of the fine structure. Another line of research tries to answer the inverse question that given certain neural responses, what the stimulus that elicits these responses is. In other words, this type of research takes the reproducibility of respnses into account and investigates how well different stimuli can be discriminated based on neural responses. (Wohlgemuth & Ronacher, 2007) have shown that different AM frequencies can be reliably discriminated (70%) based on metric distances of spike trains.

The studies reviewed so far assess response properties of cortical neurons using artificial sounds, such as frequency-modulated sounds, amplitude-modulated sounds (FM and AM tones), and ripple sounds. Such stimuli differ considerably from natural sounds, including animal vocalizations and human speech, which are not only substantially more acoustically complex but are also much more ecologically relevant. Given the theory of natural selection, one can assume that the auditory systems of species are well adapted to their specific acoustic environments. Thus, the cortical neurons are the most selectively sensitive neurons of the auditory system. In other words, natural sounds are the most efficient stimuli to activate cortical neurons. The "most efficient stimuli" theory has been explored theoretically (Lewicki, 2002; Smith & Lewicki, 2006) and experimentally (deCharms, Blake, & Merzenich, 1998; Machens, Gollisch, Kolesnikova, & Herz, 2005; Narayan, Grana, & Sen, 2006; Nelken, 2004; Woolley, Fremouw, Hsu, & Theunissen, 2005). To further assess the theory, new, quantified measurements of complex neural responses to natural sounds are needed.

1.2.2 Modulatory effects of attention on representations of features

Hierarchical organization of sound feature processing cannot alone explain several perceptual phenomena, such as auditory sensory memory, selective attention, and perceptual learning. Converging evidence from recent studies suggests that cognitive functions play an important role in modulating feature representations, a characteristic known as plasticity of the auditory system. Two kinds of plasticity, long-term (changes that last years, or even permanently) and short-term (changes that last from tens of milliseconds to minutes) exist. Here we only consider short-term plasticity, which refers to any bottom-up and top-down inputs that transiently modulate the responsiveness of the target neurons and thus also change the oscillatory properties of the local neuronal population (Jaaskelainen, Ahveninen, Belliveau, Raij, & Sams, 2007).

A typical example of bottom-up short-term plasticity is stimulus-specific adaptation (SSA), a sound-feature-specific suppression caused by a preceding auditory stimulus. The suppression is more significant when the two tones fall within the same critical band. Lateral inhibition could be the primary neurophysiologic mechanism underlying SSA (Okamoto, Ross, Kakigi, Kubo, & Pantev, 2004). Recent studies of the auditory cortex suggest that SSA plays an important role in forming auditory sensory memory and detecting novel stimuli (Jaaskelainen et al., 2004; Ulanovsky, Las, & Nelken, 2004). Auditory sensory memory, a short-term memory representation of the auditory environment, is vital for speech comprehension and auditory working memory tasks where relevant auditory information needs to be accessed quickly. Neurons in primary auditory cortex exhibit SSA at multiple resolutions, varying from hundreds of milliseconds up to tens of seconds. These multiple resolutions of SSA are postulated to combine with hierarchical feature analyses, which also occur on multiple timescales, to support adaptive behavioral functions (Ulanovsky, Las, Farkas, & Nelken, 2004; Ulanovsky, Las, & Nelken, 2004).

Recently it has been shown that SSA also plays an important role in streaming. Fishman et al. (2001) examined responses evoked by alternating frequency sequences of tones (ABAB) as a function of temporal separation between A and B, using multi-unit recordings from the auditory cortex of the awake monkey. A large temporal separation (ΔT, reciprocal of presentation rate, PR) is associated with an overall pattern of neural activity at the stimulus PR. In contrast, responses to the A and B tones are differentially suppressed when their temporal separation is small, resulting in a pattern of activity consisting predominantly of ‘A’ tone responses at half the PR. The magnitude of the suppression of ‘B’ tone response increases with spectral separation (ΔF) between A and B. The effects of ΔT and ΔF are consistent with the limitations of SSA on the temporal and spectral relationships between preceding maskers and the target. This same effect of SSA on streaming is found in songbirds (Bee & Klump, 2005).

The aforementioned studies used physically different stimuli to induce 1 or 2 stream percepts; however, stimulus-induced changes in neural response patterns confound percept-induced changes in neural activity. To dissociate these two causes and isolate the sole neural correlates of the percept, Micheyl et al. (2005) compared neural responses of the first triplet (ABA-) to the last tone within a sequence. The rationale for this lies in the ‘build-up’ phenomenon, which refers to the fact that stream segregation typically takes some time to build-up (Carlyon, Cusack, Foxton, & Robertson, 2001). In other words, the percept evoked by the first triplet (ABA-) is different from the percept induced by the last tone, although the stimuli are unchanging. By using this approach, a correspondence between the temporal dynamics of neural responses in awake monkeys and the perceptual build-up of stream segregation as measured in humans was revealed, reflecting a potential association between cognitive processes and neural patterns.

Neural patterns are also modulated by other cognitive functions, such as attention (J. Fritz, Shamma, Elhilali, & Klein, 2003; J. B. Fritz, Elhilali, & Shamma, 2005; Polley, Steinberg, & Merzenich, 2006), and behavioral context (Durif, Jouffrais, & Rouiller, 2003; Eliades & Wang, 2005; Wang, Lu, Snider, & Liang, 2005). The present studies focus on attention, which is a powerful mechanism that enables us to focus on a small subset of the information contained in a busy scene. Attention, as an expression of cognitive control over perception, can be allocated to a particular space and/or a particular feature of sound, known as space-based attention and feature-based attention, respectively. In addition, the deployment of attention can be either goal-directed and voluntary, or stimulus-driven and involuntary. The following two studies show plasticity effects of attention based on spectrotemporal features. Fritz et al. (2003) found that when ferrets were trained to attend to a target note, the spectral aspect of the spectrotemporal receptive fields (STRF) of cortical neurons was transiently changed to encompass the frequency of the target tone, while the temporal aspect of the STRF of cortical neurons was changed to encompass the temporal range of the target tone when the animals were trained to attend to temporal sound features. A similar task-specific reorganization of frequency and intensity maps in auditory cortex was observed in rats by (Polley, Steinberg, & Merzenich, 2006). Results from these studies suggest that feature maps are transiently modulated by top-down processes.

Studies of representational plasticity in auditory cortex of animals have largely focused on the spectral feature. Few studies have examined representational plasticity of the temporal modulation feature. Kilgard & Merzenich (1998a; 1998b) found that pairing of nucleus basalis stimulation with sounds changes the characteristics of tMTFs in the rat primary auditory cortex. Pairing NB stimulation with slowly modulated stimuli markedly decreased the cortical response to rapidly presented stimuli, whereas pairing with fast modulated stimuli significantly increased the maximum cortical following rate. These results indicate that processing temporal information in the auditory cortex undergo plastic reorganization, and may be modified by various auditory tasks.

1.3 Large-scale neurophysiology: MEG/EEG studies

Magnetoencephalography (MEG) is a non-invasive technique that measures large-scale neural activity via its electromagnetic signature (Hamalainen, Hari, Ilmoniemi, Knuutila, & Lounasmaa, 1993); the cortical neuron populations activated by appropriate stimuli carry ionic currents which emanate external magnetic fields. The magnetic field, which passes transparently through brain tissue and the skull, is picked up by superconducting quantum interference devices (SQUIDs) in the MEG sensors. MEG provides moderate spatial resolution (5-10 mm) and excellent temporal resolution (1 ms). Its high temporal resolution makes it a powerful tool to explore the timing of neural processes at the level of cell assemblies (Baillet & Garnero, 1997; Lutkenhoner et al., 2003). Electroencephalography (EEG) has the same temporal resolution as MEG, but is subject to extreme distortion as it propagates through the brain tissue and the scalp.

Over the past decade, these two fruitful brain imaging techniques have provided insight into neural mechanisms at the systems level. MEG/EEG data has been examined mainly in three domains: time, spectral, and source domains. Neural responses unfold along the time dimension. Thus a time-waveform is a natural measurement of neural activities. The MEG/EEG waveform may be characterized by an event-related potential (ERP), which is derived by averaging recordings across a large number of trials. The underlying assumption is that responses to repeated stimuli are phase-locked with each other and can be recovered from averaging, while noise might be cancelled out by averaging due to its random phase across trials. The most intensively researched ERP component is the M100, a deflection observed approximately 100 milliseconds after stimulus onset. The M100 amplitude and M100 latency depend on both acoustic features--for example, on the intensity and frequency of acoustic stimuli (Roberts, Ferrari, Stufflebeam, & Poeppel, 2000)--and stimulus and behavioral contexts such as stimulus repetition (Budd, Barry, Gordon, Rennie, & Michie, 1998), long-term training/experience (Shahin, Bosnyak, Trainor, & Roberts, 2003), and selective attention (Fujiwara, Nagamine, Imai, Tanaka, & Shibasaki, 1998; Hillyard, 1998; Hillyard, Hink, Schwent, & Picton, 1973; Poeppel et al., 1996). Another potentially important component of the ERP is the mismatch negativity (MMN), elicited by infrequent changes along a particular feature dimension in the auditory stimulus sequence. The elicitation of MMN, typically observed between 100ms and 200ms after stimulus onset, is independent of attention (Naatanen, 1992). The MMN is assumed to represent an early process of deviance detection based on an auditory memory trace, and therefore it is related to SSA (D. R. Moore, 2003; Sussman & Steinschneider, 2006; Winkler, Karmos, & Naatanen, 1996).

It has been generally held that an ERP is the result of a set of stimulus-evoked, fixed-latency, fixed-polarity brain events. But recently it has been shown that ERP might be generated by stimulus-induced phase-resetting in ongoing brain oscillations, such as the delta, theta, and alpha frequency bands (Figure 1.4) (Gruber, Klimesch, Sauseng, & Doppelmayr, 2005; Makeig et al., 2002; Penny, Kiebel, Kilner, & Rugg, 2002). This new perspective conforms to the modern trend to interpret MEG data in the spectral domain. Distinct mental states have been associated with neural activity at different frequencies. For example, strong neural activity in the theta band (4 Hz – 8 Hz) is related to drowsiness in subjects, while activity in the beta range (14 Hz – 30 Hz) is associated with a highly alert and focused mental state (see review, da Silva, 1991). Neural activity at the modulation rate of stimuli, known as the auditory steady state response (aSSR), is regarded as the representation of the temporal modulation feature at the systems level. This will be discussed in detail in section 1.3.1.

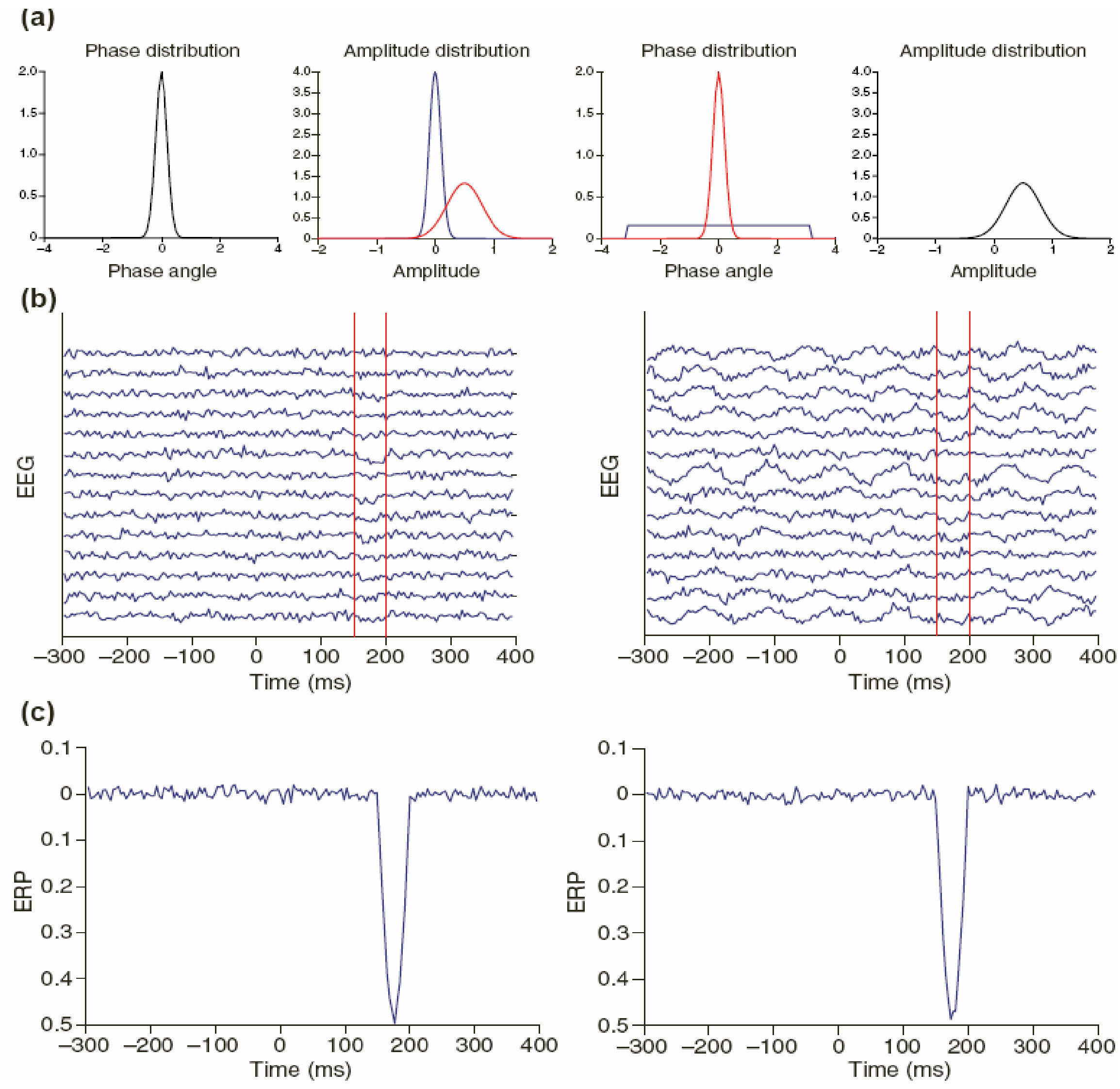

Figure 1.4: How an amplitude-modulated (AM) electroencephalogram (EEG) and phase-modulated (PM)-EEG can give rise to the same event-related potential (ERP). (a) In each trial and in each period

(background or stimulus-induced), the amplitude and phase of each sinusoid are drawn from their respective distributions. Background distributions are shown in blue and stimulus-induced distributions in red. Black indicates the same distribution for both periods. (b) Synthetic data from 15 trials of EEG at a single electrode. On each trial, data was generated by adding Gaussian noise onto 10-Hz sinusoids. Within each trial there is a stimulus-induced period (150–200 ms post-stimulus, between the red lines), the rest of the trial constituting a background period. (c) The ERP computed from 1000 trials of such data. The AM and PM ERPs are identical (to within the noise limit). This figure is adopted from (Penny, Kiebel, Kilner, & Rugg, 2002).

To link the MEG/EEG responses with underlying neural activity, MEG/EEG recordings are projected back to the source domain. The underlying assumption is that MEG/EEG records the superimposed magnetic fields generated by several underlying neural sources. Several methods have been proposed to recover the neural source signal. However, the problem of source decomposition is an inverse problem with non-unique solutions, resulting from the fact that the same magnetic/electronic fields could be generated by different configurations of neural sources (Baillet et al., 2001). Source analysis of the M100 places its origin in the superior temporal gyrus (Lutkenhoner et al., 2003). The contour maps of several ERP components, aSSRs, and neural source signals are shown in Figure 1.5 A-E.

Note that the analysis and interpretation of MEG/EEG responses in the three domains are not mutually exclusive but might relate to each other. For example, the M100 might be considered as a superposition of oscillations of the alpha and delta frequency bands (Gruber, Klimesch, Sauseng, & Doppelmayr, 2005), reflecting a transformation from the time to the spectral domain. In contrast, the 40 Hz aSSR might be formed by repetitive middle latency responses (MLR), hence representing a transfer from the spectral to the time domain.

The discussion in sections 1.3.1 and 1.3.2 parallels sections 1.2.1 and 1.2.2. Representations of temporal modulation and the modulatory effects of attention are discussed, but from the perspective of the neural ensemble instead of the single neuron.

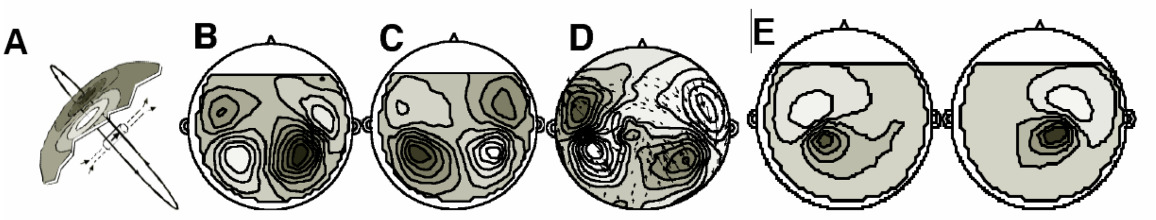

Figure 1.5: Cartoon illustration of a compact neural current source generating a source/sink pair of magnetic field responses. B-E. Spatial field distribution of magnetic responses, superimposed on a "flattened" head. Dark contours represent magnetic flux out of the head, light contours into the head. These examples show data from individuals, using different paradigms. They share common features despite their very different underlying responses. All the responses reflect a compact dipolar neural source beneath the location of the steepest gradient. B. Spatial distribution of the auditory evoked M50 response (~50 ms post onset) to a broadband noise stimulus. The entire spatial distribution can be parsimoniously ascribed to a pair of compact dipolar sources, one in each auditory cortex. C. Spatial distribution of the auditory evoked M100 response (~100 ms post onset) to the same broadband noise. Again, a compact neural source is found in each auditory cortex, but with a polarity opposite to the M50. D. Spatial distribution of an auditory steady state response (SSR, details below). Unlike the M50 and M100 examples, this is not the spatial distribution at a particular time, but at a particular frequency. Nevertheless, the pattern is strikingly similar: separate neural responses in each hemisphere arise from compact sources in each auditory cortex. E. Spatial distributions determined by independent component analysis (ICA). The responses generated by the left and right cortices are independent and so appear as separate response components. Unlike the previous responses, these are not spatial distributions at any one time or frequency, but rather reflect the spatial

distribution of a single component (compact source) over the entire duration of the response (they are also not auditory—the sources are superior to the temporal lobe)

1.3.1 Representations of temporal modulation

Neural responses in human cortex to modulated stimuli can be characterized by the auditory steady state response (aSSR), an evoked response that has the same frequency as the stimulus modulation rate. SSR was originally measured in the visual domain (D. Regan, 1966) and was observed in the auditory realm later (Campbell, Atkinson, Francis, & Green, 1977). Since then, the SSR to auditory stimuli (aSSR) has been studied by using EEG/MEG (Galambos, Makeig, & Talmachoff, 1981; Hari, Hamalainen, & Joutsiniemi, 1989; Picton, John, Dimitrijevic, & Purcell, 2003; Rees, Green, & Kay, 1986; Ross, Borgmann, Draganova, Roberts, & Pantev, 2000). Much evidence has shown that the amplitude of the aSSR decreases with increasing modulation rates, with a local peak around 40 Hz (Rees, Green, & Kay, 1986; Ross, Borgmann, Draganova, Roberts, & Pantev, 2000). In addition, cortical and subcorical sources for aSSR at low and high rates, are suggested (Kuwada, Batra, & Maher, 1986; Pantev, Roberts, Elbert, Ro, & Wienbruch, 1996; Ross, Borgmann, Draganova, Roberts, & Pantev, 2000).

In a recent study by Liegeois-Chauvel et al. (2004), phase-locked neural responses to SAM noises with modulation rates between 4 and 128 Hz were recorded in four auditory cortical areas with intracerebral electrodes. The study shows that the auditory cortex is responsive to modulating rates below 50 Hz. Furthermore, for these low modulation rates, a number of MTFs are band-pass in shape, demonstrating a selective encoding of AM rates in the auditory cortex. This result provides physiological support for the hypothesis that the human auditory system decomposes the temporal envelope of sounds into its constituent AM components. Moreover, the results show a predominant response of cortical auditory areas to the lowest AM frequencies (4–16 Hz). This range matches the range of AM rates crucial for speech intelligibility, therefore emphasizing the role played by these initial stations of cortical processing in the analysis of speech. It is noteworthy that both the cut-off rate and bandpass shape are consistent with findings in single-neuron recording.

Due to the non-linearity of the auditory system, processes of sounds with multiple modulations cannot be reduced to a superimposition of the responses to single modulations. Thus, neural processes of multiple modulations provide independent evidence of this non-linearity function. Neural processes of sounds with multiple modulations are still controversial. Using multiple SAM stimuli with rates ranging from 80 Hz to 105 Hz, Lins and Picton (1995) showed superimposed aSSR and found no interaction between modulation rates. However, significant interactions are observed using modulation rates from 35 Hz to 55 Hz (Draganova, Ross, Borgmann, & Pantev, 2002; John, Lins, Boucher, & Picton, 1998). Our first study provides a systematic investigation of the encoding of multiple modulation rates at a low range (< 25 Hz), the modulation band that is critical for speech recognition.

Two modulation modulation-filter theories have been proposed to account for the neuronal physiology and behavioral findings. Viemeister (1979) postulated that the auditory cortex functions as a low-pass modulation filter, based on the observation that subjects' sensitivity for detecting amplitude modulation decreases with the increase of modulation rates. In an alternative modulation-filter theory, Dau et al. (1997) argue that different groups of neurons tune to distinct modulation rates, forming a modulation filter bank. The result of our first study reconciles these two models.

1.3.2 Modulatory effects of attention on representations of features

Much evidence has shown that the cortical response, measured by MEG/EEG, is modulated by stimulus context, behavioral context, and cognitive processes. For example, the power of the induced gamma-band activity (Singer & Strategies, 1999; Tallon-Baudry & Bertrand, 1999) and the alpha rhythm (Vanni, Revonsuo, & Hari, 1997) have been associated with the mental construction of a coherent object, suggesting an important role of gamma and alpha oscillations in binding the distributed components of an object. Recently, Luo and Poeppel (2007) found that phase of theta oscillations is modulated by speech-specific features. The phase pattern at 4-8 Hz potentially discriminates the corresponding spoken sentence. The remainder of this section considers the modulatory effect of attention.

Starting with their pioneering work in the early 1970s, Hillyard et al. (1973; 1998) have used EEG to investigate the neural correlates of the control of spatial attention in a dichotic paradigm, where subjects listen selectively to a series of tone pips in one ear and ignore concurrent tone pips in the other ear. The amplitude of N1, a counterpart of the M100 in EEG, was substantially larger for the attended tones, reflecting transient modulation of activity in the neural populations involved in the obligatory sensory analysis of acoustic stimuli in auditory cortex. Subsequently, using a similar dichotic-listening protocol, selective attention has been found to enhance the neural response at an earlier stage (20-50 ms) (Woldorff et al., 1993) and at the gamma-band (40 Hz) (Tiitinen et al., 1993), suggesting a gain-based model of selective attention.

An alternative mechanism, a sharpening-based model, was recently explored in a dichotic protocol by simultaneously presenting a tone target with a notched noise masker (Kauramaki, Jaaskelainen, & Sams, 2007; Okamoto, Stracke, Wolters, Schmael, & Pantev, 2007). During the selective attention conditions, the suppressant effect of the noise notch width on the M100 was decreased, but as a function significantly different from the multiplicative function one would expect on the basis of the gain model. This suggests that auditory selective attention cannot be fully explained by a gain model, but rather that selective attention additionally enhances the frequency selectivity of the auditory cortex.

The investigation of attention has focused primarily on space-based attention. Growing behavioral and neurophysiological evidence has shown, however, that selective attention frequently operates on feature dimensions. Components in an auditory scene are organized and attended according to a particular feature. In an electrocorticography study by Bidet-Caulet et al. (2007), a stream segregation paradigm was used to investigate feature-based (here, temporal modulation) top-down attention in epileptic patients. Sound onset asynchrony was manipulated to induce the segregation of two concurrent auditory streams, which consisted of amplitude-modulated tones at different carrier and modulation rates. Patients were presented with the stimuli while they either performed an auditory distracting task or actively selected one of the two concurrent streams. Selective attention was found to affect steady-state responses in the primary auditory cortex, and transient and sustained evoked responses in secondary auditory areas.

Recently, neuromagnetic correlates of streaming were investigated in human auditory cortex. Gutschalk et al. (2005) demonstrated a strong correlation between the percept of streams and the amplitude of two ERP components (M50 and M100), which may provide a macroscopic expression of the SSA in a single neuron. Snyder et al. (2006) investigated the effects of attention on auditory stream segregation by comparing neuromagnetic responses in two tasks, where participants either attended to sound stimuli or watched a muted movie while being presented identical stimuli. An enhancement of the auditory-evoked response (P1–N1–P2 and N1c) to the B tone is found in the attention condition. In addition, the N1c enhancements were larger at right than left temporal electrodes, suggesting a right-hemisphere dominance for stream segregation.

1.4 fMRI research

Functional magnetic resonance imaging, or fMRI, is a technique for measuring activity in the brain by detecting the changes in blood oxygenation and flow that occur in response to neural activity – when a brain area is more active it consumes more oxygen, and to meet this increased demand blood flow increases to the active area. fMRI has been used to produce activation maps showing which parts of the brain are involved in processing particular perceptual features or cognitive process.

In fMRI paradigms, subjects are engaged in two or more mental states, by being given different stimuli or tasks in blocks, and then the activation maps in different mental states are compared. Those brain regions with significant changes are associated with particular feature extraction or cognitive processes. fMRI data has also been studied using multivariate analysis, i.e., both the pattern of activation at all voxels and the interaction between these areas are deemed as information carriers and are examined. A main deficiency of the hemodynamic method is its coarse temporal resolution (~1 second). Hence, fMRI fails to accurately measure temporal activity and is insufficient to probe the temporal information encoded in the representations of temporal modulation.

An important feature of cortical neural processes revealed by fMRI studies is hemispheric asymmetry, the lateralization of function found in the auditory cortex. The most extensively investigated lateralization is speech/music lateralization, which states that language processing occurs mainly in the left hemisphere while music processing happens in the right hemisphere. One prevalent hypothesis proposes that such functional lateralization arises from differences in the early spectrotemporal computations performed in auditory cortices that transform sensory representations of signals into more abstract perceptual codes (Zatorre & Belin, 2001). More specifically, temporal features are predominantly processed in the left hemisphere and spectral features in the right hemisphere. An alterative theory, the asymmetrical sampling hypothesis, proposes that both hemispheres analyze auditory signals on two time scales (one corresponding to ~25–50 ms and the other corresponding to ~200–300 ms), with different principal temporal scales in each hemisphere. Information processed on longer timescales (dominant in music) is associated dominantly with right hemisphere areas, whereas information processed on a shorter timescale (dominant in speech, such as formant transitions) primarily projects to the left areas (Poeppel, 2003).

Lateralization is also found for high-level cognitive functions. Haynes and Rees (2005) found that distinct distributed fMRI response patterns are associated with the dominance of different conscious percepts of a visual object. It was also shown that the left hemisphere is largely responsible for attending to relevant stimuli, while the right hemisphere suppresses irrelevant signals (Alcaini, Giard, Echallier, & Pernier, 1995). The following section reviews the fMRI research on cortical representations of temporal modulation.

1.4.1 Representations of temporal modulation

To investigate the responsiveness of the auditory cortex to temporal modulation, (Hart, Palmer, & Hall, 2003) compared fMRI responses to SAM, SFM, and un-modulated stimuli. Relative to a silent baseline, the modulated tones activate widespread regions of the auditory cortex bilaterally along the supra-temporal plane. Compared to the un-modulated tones, both AM and FM tones generate significantly greater activation in lateral HG and the planum temporale. Largely overlapping activation between AM and FM indicates a common neural code to both AM and FM in the auditory cortex.

A parametric study of SAM was provided by Giraud et al. (2000) by using noise bursts with different repetition rates, ranging from 4 to 256 Hz. This range includes low AM rates (up to 32 Hz), essential for the perception of articulation and the syllable, and high AM frequencies (above 64 Hz), essential for the perception of pitch and prosody. The results suggested a hierarchically organized filter bank, with each processing level responding preferentially to a given AM frequency: 256 Hz for the lower brainstem, 32-

256 Hz for the inferior colliculus, 16 Hz for the medial geniculate body, 8 Hz for the primary auditory cortex, and 4-8 Hz for secondary regions.

In addition, the shape of the response wave in the auditory cortex changes dramatically with increases in the stimulus rate; low rates elicit a sustained response, whereas high rates elicit an unusual phasic response. The spatial representation of AM rate varies according to the response type. Sustained responses yield maps of low frequencies organized in large clusters, while transient responses yield maps of high frequencies represented by a mosaic of small clusters. It was suggested that the transition from a sustained to a phase response shape may be correlated with the perceptual shift from burst trains to a fused continuous auditory event (Harms & Melcher, 2002).

Enhancement of auditory-cortex responses to attended versus non-attended stimuli has been recently investigated using fMRI (Ahveninen et al., 2006). Rather than there being a single, unitary, selective-attention effect, the task given to the subject seems to determine which parts of the auditory cortex are modulated by top-down inputs. The anterior auditory cortex (the putative ‘what’ processing pathway) exhibited enhanced selectivity to phonemes when subjects directed their attention to phonetic features. Conversely, selectivity of the posterior auditory cortex (the putative ‘where’ processing pathway) to spatial location of auditory stimuli was enhanced when subjects directed their attention to stimulus locations. This dissociation of ‘what’ and ‘where’ pathways are consistent with findings in the visual domain (Ungerleider & Mishkin, 1982).

1.5 Summary

ASA, the process of organizing acoustic input from multiple sound sources into meaningful auditory objects, and speech perception are two advanced cognitive functions of the auditory system. Emerging evidence suggests two principles underlying these cognitive functions, a hierarchical organization of sound feature processing and the plasticity of the organization. Ascending the neural-axis of the auditory system, neurons extract increasingly complex sound features, ranging from frequency, to location, to temporal modulation. High-level cognitive functions, such as attention, emotion, and life experience, affect our percept by shaping the neural representations of features. This chapter reviewed the present knowledge of the neural correlates of these two principles. We limited the focus of this survey to the encoding of temporal modulation and the effect of attention on the encoding mechanisms.

Neurophysiology, large-scale neurophysiology, and fMRI studies have yielded a great deal of information about the representations of single temporal modulations. Three basic principles are emerging from this body of work. First, ascending the auditory axis, the highest valid modulation rate decreases, ending with upper cut-off frequencies around 50 Hz in auditory cortex. Second, in auditory cortex, slow modulation is encoded by synchronized neuron populations using temporal coding, while relatively fast modulation is encoded by non-synchronized neuron populations using rate coding. Last, a high percentage of temporal-coding neurons show bandpass characteristics; thus a neuron population forms a periodotopy.

Based on these findings, two further questions arise. First, what are the neural representations of multiple temporal modulations? These modulations are an essential feature of speech but have been inadequately investigated. Considering the non-linearity of the auditory cortex, it might not be a superimposition of responses to single modulations. Second, does attention to temporal modulation shape the representations? The present studies address these two questions by using MEG. Chapter 2 discusses representations of multiple temporal modulations, an essential feature of vocalization and speech that has been inadequately investigated. Chapter 3 and chapter 4 explore the effects of attention on cortical representations of temporal modulation using ‘informational masking’ and ‘stream segregation’ paradigms, respectively. In the former paradigm, the target stream is embedded by maskers, while in the latter case two streams are presented concurrently, forming a sound rivalry situation.

Chapter 2: Neural representations of compound modulations

2.1 Introduction

Natural sounds, including animal vocalizations and human speech, are time varying signals. Information contained in the dynamic temporal structure is crucial for content identification and communication. The presentation of a dynamic temporal pattern in only a few broad spectral regions is sufficient for the recognition of speech; the most critical speech content information is preserved in the slow temporal envelope (below 20 Hz) (Drullman, Festen, & Plomp, 1994; Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995). Temporal modulation is also an important cue for pitch processing (B. C. J. Moore & Gockel, 2002), stream segregation (Bregman, 1990), and signal detection in the presence of competing maskers (Hall & Harvey, 1984).

To investigate auditory encoding of temporal information in complex sounds, sinusoidally amplitude modulated (SAM) stimuli have been widely used in physiological studies. The neural physiological responses to SAM stimuli along auditory neuron-axis have been extensively investigated (for a review, see Joris, Schreiner, & Rees, 2004). The available evidence suggests that ascending the auditory axis, the highest responsive modulation rate decreases, ending with upper cut-off frequencies around 50 Hz in the auditory cortex (Bieser & Muller-Preuss, 1996; Eggermont, 1998; Schreiner & Urbas, 1988). Moreover, slow modulation information (< 20 Hz) is encoded in the auditory cortex by synchronized neuron populations, while relatively fast modulation (20 Hz < f < 50 Hz) is encoded by non-synchronized neuron populations using rate coding (Liang, Lu, & Wang, 2002; Wang, Lu, & Liang, 2003). A high percentage of temporal-coding neuron shows band pass characteristics, i.e. neurons tune to distinct modulation rates (Gaese & Ostwald, 1995; Lu, Liang, & Wang, 2001).

Neural responses to modulated stimuli in human sensory cortex have been examined by electroencephalography (EEG) and magnetoencephalography (MEG), powerful tools to explore the timing of neural processes at the level of cell assemblies (Baillet & Garnero, 1997; Lutkenhoner et al., 2003). The EEG/MEG signals evoked by modulated sounds are characterized by the auditory steady state response (aSSR), a spectral component at the same frequency as the stimulus modulation frequency (Campbell, Atkinson, Francis, & Green, 1977; Galambos, Makeig, & Talmachoff, 1981; Hari, Hamalainen, & Joutsiniemi, 1989; Picton, John, Dimitrijevic, & Purcell, 2003; Rees, Green, & Kay, 1986; Ross, Borgmann, Draganova, Roberts, & Pantev, 2000). Much evidence has shown that amplitude of aSSR decreases with increasing modulation rates, but with a local peak around 40 Hz (Rees, Green, & Kay, 1986; Ross, Borgmann, Draganova, Roberts, & Pantev, 2000).

Current literature focuses on the neural processing of single modulated stimuli. However, in typical situations, e.g. in communications, sounds contain concurrent modulations. Hence stimuli with multiple modulations are a more realistic approximation of natural sounds. Neural responses to compound modulated stimuli were first examined by Regan and Heron (1969) using visual stimuli. Four visual stimuli were modulated at different rates and presented to one quadrant of the visual field. Neural responses to individual stimulus could then be recognized by distinct spectral components corresponding to the modulation frequency. Afterwards, neural responses to auditory stimuli with multiple modulations were examined by several studies (Draganova, Ross, Borgmann, & Pantev, 2002; John, Lins, Boucher, & Picton, 1998; Lins & Picton, 1995; M. P. Regan & Regan, 1988). These studies establish that different modulation rates, applying on the same carrier, can be dissociated by corresponding spectral components in neural responses. However most of these studies focused on frequencies above 40 Hz, which are far above the modulation rates critical for speech. The present study investigates the auditory encoding of modulations below 25 Hz, the critical modulation band for speech recognition.

Two modulation-filter theories have been proposed to explain the neurophysiology and behavioral findings. The first proposed that the demodulation of input signals is induced by the half-wave rectification and compressive processes occurring in the periphery. A low-pass filter in subsequent auditory stages is also incorporated to account for the observation that subjects' threshold for detecting modulation decreases with increased modulation rates (Viemeister, 1979). The second scheme models the detection of modulation by a bank of band-pass filters that are located in the central stage and are sensitive to different ranges of modulation frequency (Dau, Kollmeier, & Kohlrausch, 1997).

The present study addresses two questions. First, how are concurrent amplitude modulations represented in the auditory cortex? Secondly, how do the neural representations of concurrent modulations fit into the current debate on modulation filters? We employ SAM stimuli containing both single and compound modulations with varied spectral distance in a MEG study. The working hypothesis is that if the auditory cortex works as a single modulation filter, given a stimuli with distant compound modulations (for example, 4 & 21 Hz), the output consists of responses at both modulation and interaction rates, due to the non-linearity of processing in the auditory cortex. Alternatively, if the auditory cortex works as a modulation filter bank with limited bandwidth, distant compound modulations activate distinct filters, no interaction responses are evoked. Thus the interaction responses can be used to evaluate the two models. Furthermore, the argument can be applied to measure the bandwidth of each filter, if the modulation filer bank model is justified. For example, to measure the bandwidth of filter with 20 Hz center rate, a stimulus with near compound modulations (for example, 18 & 21 Hz) is employed. If interaction responses are observed, the concurrent modulations are processed by the same filter. Otherwise, they are processed by different filters, suggesting that the bandwidth of the 20 Hz filter is less than 3 Hz (21 – 18 Hz).

2.2 Methods

2.2.1 Subjects

Sixteen subjects (7 males; mean age 24 years) participated in this MEG study. All subjects were right handed (Oldfield, 1971), had normal hearing and no history of neurological disorder. The experiments were approved by the University of Maryland Institutional Review Board and written informed consent was obtained from each participant. Subjects were paid for their participation.

2.2.2 Stimulus design

The stimuli were generated using MATLAB (MathWorks Inc, Natick MA). Sounds were 50.25 seconds in duration with a 15 ms onset and offset cosine ramps and sampled at 44.1 kHz. Three types of conditions were employed: the single AM condition (the envelope of the stimuli contains just a single frequency component f_{am1}, the near AM-AM condition (the envelope of the stimuli consists of two frequency components f_{am1} and f_{am2}, with 3 Hz as the rate separation) and the distant AM-AM condition (the envelope of the stimuli consisted of two frequency components f_{am1} and f_{am2} with 17 Hz as the rate separation). The stimuli were created according to Equation 2.1.

$$y(t) = A\left\{2 - m\left[\cos(f_{am1}t) + \cos f_{am2}t)\right]\right\} \times \sin 2\pi f_c t \qquad \text{(Eq. 2.1)}$$

Six single AM stimuli were generated with modulation frequencies of 4, 5, 18, 19, 21 and 22 Hz. Two distant AM-AM stimuli were created by combining 4 & 21 Hz and 5 Hz & 22 Hz, respectively. Two near AM-AM stimuli were made with 18 & 21 Hz and 19 & 22 Hz. Modulation rates around 4 Hz characterize syllabic rhythms of speech, while rates around 20 Hz characterize phoneme segmentation. Finally, these ten envelopes were each applied to two different carriers: a pure tone at 707 Hz and a 5 octave pink noise centered at 707 Hz, giving a total of 20 stimuli. The envelope and spectra of three example stimuli from each condition are illustrated in Figure2.1

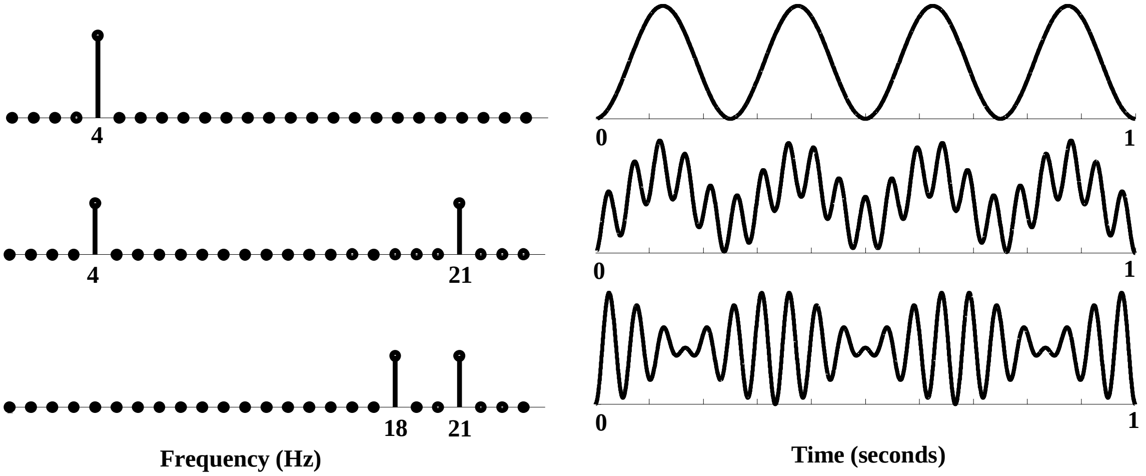

Figure 2.1: Stimulus description. a) Spectrum of typical stimulus envelopes of each condition. The top one, a single AM sound, modulation rate 4 Hz; the middle one, a near AM-AM sound, modulated at 18 & 21 Hz; the bottom one, a distant AM-AM sound, modulated at 4 & 21 Hz. b) The corresponding temporal waveform envelope. The carrier can be tone or noise.

2.2.3 Experimental procedure

Subjects were placed horizontally in a dimly lit magnetically shielded room (Yokogawa Electric Corporation, Tokyo, Japan). Stimuli were presented using Presentation software (Neurobehavioral Systems, Albany, CA). Sounds were delivered to the subjects' ears with 50 Ω sound tubing (E-A-RTONE 3A, Etymotic Research, Inc), attached to E-A-RLINK foam plugs inserted into the ear-canal and presented binaurally at a comfortable loudness of approximately 70 dB SPL.

Before the main experiment, a pre-experiment was run, where a 1 kHz, 50 ms tone pip was presented about 200 times. The inter-stimulus intervals (ISIs) were randomized between 750 ms and 1550 ms and subjects were instructed to count the total number of the tone pips. The aim of this task was to obtain the M100 response (a prominent peak approximately 100 ms after pip onset), which was used to verify that the location and strength of neural signals fell in a normal range.

In the main experiment, subjects listened passively to the acoustic stimuli while MEG recordings were taken. Each stimulus was presented once. ISI were randomized between 1800 to 2200 ms. MEG recordings were conducted using a 160-channel axial gradiometer whole-head system (Kanazawa Institute of Technology, Kanazawa, Japan). Its detection coils are arranged in a uniform array on a helmet-shaped surface of the bottom of the dewar, with about 25 mm between the centers of two adjacent 15.5 mm diameter coils. Sensors are configured as first order axial gradiometers with a baseline of 50 mm; their field sensitivities are 5 fT/√Hz or better in the white noise region. Three of the 160 channels are magnetometers separated from the others and are used as reference channels in noise filtering methods. The magnetic signals were bandpassed between 1 Hz and 200 Hz, notch filtered at 60 Hz, and sampled at the rate of 500 Hz. All 157 neural channels were denoised twice with a Block Least Mean Square (LMS) adaptive filter: first using the 3 external reference channels (Ahmar & Simon, 2005), and secondly using the 2 channels with the strongest cardiac artifacts (Ahmar, Wang, & Simon, 2005).

2.2.4 Data analysis

For each stimulus, an epoch of duration 50 seconds was extracted from 0.25 seconds post-stimulus to the end of the stimulus. Each single trial response was transformed to a complex frequency response (of 0.02 Hz resolution and 250 Hz extent) by the discrete Fourier Transform (DFT). The neural responses at 6 modulation frequencies (*4, 5, 18, 19, 21, 22 Hz*) and 6 potential interaction frequencies (*3, 17, 25, 27,* *39, 41 Hz*) were obtained for each stimuli and each channel. The bulk of the analysis was based on the normalized neural responses, defined to be the squared magnitude of the spectral component at the target frequency divided by the average squared magnitude of the spectral components ranging from 1 Hz below to 1 Hz above the target frequency (excluding the component at the target frequency), averaged over the 20 channels with the strongest normalized neural responses.

The spatial pattern of the neural responses was represented by a phasor map, a graph of the complex (magnitude and phase) magnetic field on all channels. For each channel, the length of the vector arrow is proportional to the magnitude of the investigated frequency and the direction of the arrow represents the phase according to standard polar coordinates. Red and green contours represent the magnetic field strength projected onto the line of constant phase that maximizes the projected field's variance (Simon & Wang, 2005).

To investigate whether adding a second modulation change the neural responses to the original modulation reduction level (RL) is defined to be difference in the normalized neural responses at modulation frequencies evoked by single and by compound modulated stimuli, normalized by responses to single modulated stimuli (Eq. 2.2). For example, RL at 4 Hz was calculated by comparing the normalized neural response at 4 Hz evoked by a compound (4 & 21 Hz) modulated sound with response at 4 Hz evoked by a single (4 Hz) modulated sound, then dividing the latter term.

$$RL = \frac{R_s - R_c}{R_s} \qquad \text{(Eq. 2.2)}$$

A reduction value of 0 indicates that adding additional modulation does not affect the neural response to the original modulation, whereas a reduction value of 1 indicates adding additional modulation eliminate the neural response to the original modulation. The reduction levels were obtained for each of the 6 modulation frequencies and each bandwidth.

To investigate the non-linearity of the cortical representations of modulations, interaction level (IL) was defined to be difference in the normalized neural responses at potential interaction frequencies evoked by associated and non-associated stimuli (Eq. 2.3). Associated stimuli are compound modulated sounds whose envelop has a distortion component as the potential interaction frequency. Hence neural activities at the interaction frequency are expected. The rest stimuli are grouped as non-associated stimuli, whose neural responses provide a baseline to examine the extent of interaction activities. For example, IL at 3 Hz was calculated by comparing the mean normalized neural response at 3 Hz evoked by two associated sounds compound (18 & 21 Hz, 19 & 22 Hz, respectively) modulated sound with the mean normalized neural response to the rest of the stimuli.

$$IL = R_a - R_{nona} \qquad \text{(Eq. 2.3)}$$

Potential Interaction frequencies were furthered labeled as three categories: difference frequencies (3 *Hz,* 21 – 18 / 22 - 19; *17 Hz*, 21 – 4 / 22 – 5), combination1 frequencies (*39 Hz*, 21 + 18; *25 Hz*, 21 + 4*)* and combination2 frequencies (*41 Hz*, 19 + 22; *27 Hz*, 22 + 5). Note that different frequencies may have the same label for near and distant AM-AM conditions. IL was averaged within each category and each modulation condition (near vs. distant).

2.3 Results

2.3.1 Neural responses at modulation frequencies

Using the high temporal resolution of MEG, we measure the neural responses to single and compound modulated sounds in a passively listening paradigm. Neural responses at modulation frequencies are detectable regardless of whether the modulation are presented alone or accompanied by an additional modulation. Figure 2.2a shows neural responses to a compound modulated stimulus (a broadband noise modulated by 4 and 21 Hz concurrently) of a typical subject. Neural responses of compound modulations are dissociated from each other, emerging as prominent peaks at individual modulation rates in the neural signal. This differential activation is suggestive evidence of the modulation filters operating at the central auditory stage. In addition, the MEG magnetic field distributions of neural responses to individual modulations, examples of which are shown in Figure 2.2b, reveal the stereotypical pattern for neural activity originating separately in left and right auditory cortex.

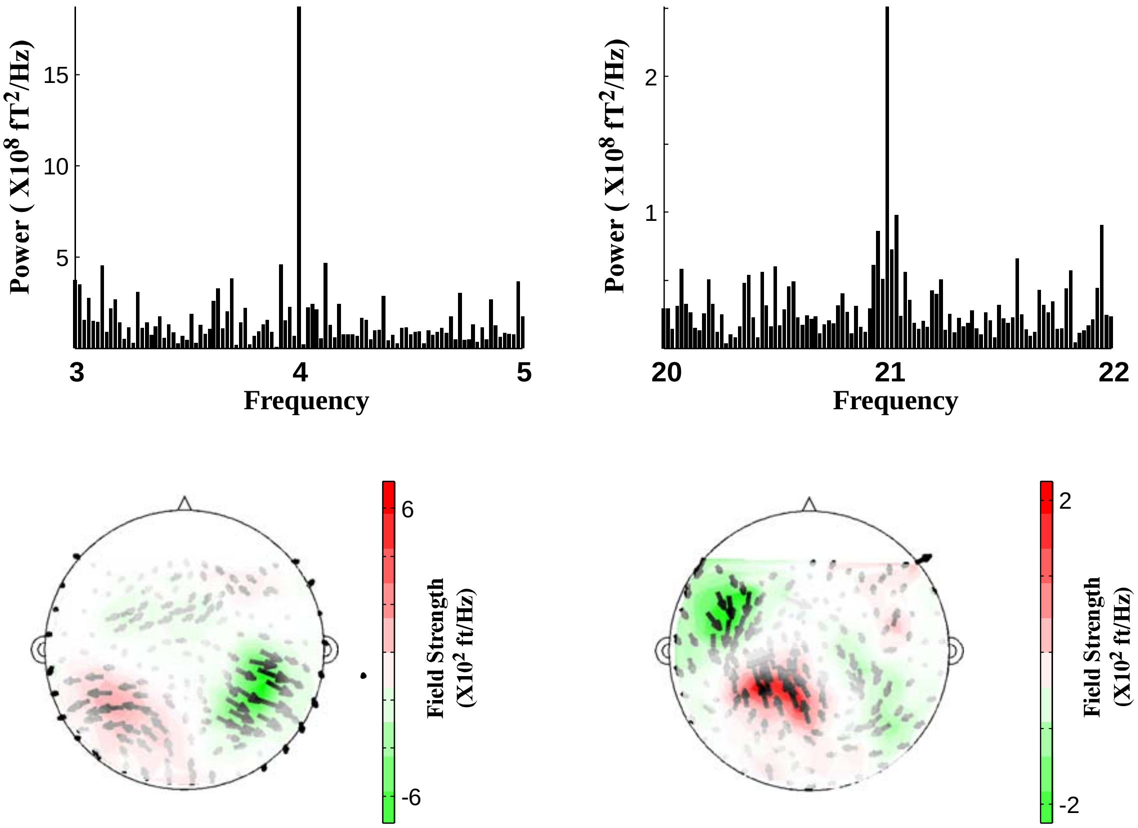

Figure 2.2: a) Power spectral density of MEG responses for a single subject, averaged over twenty channels. The stimulus is a broadband noise modulated by 4 Hz and 21 Hz concurrently. b) The MEG magnetic field distributions of the neural response to modulations. The SSR at each significant channel is represented by an arrow whose length is proportional to the magnitude of the SSR and whose direction represents the phase. Red and green contours represent the magnetic field strength projected onto a line with the constant phase. Each hemisphere is dominated by a dipolar pattern.

Although compound modulated stimuli evoke neural responses at individual modulations, the power of neural responses is reduced compared to when the modulation is presented alone. The reduction is characterized by reduction level, defined as difference in neural responses at modulation frequencies between single and compound modulations and then normalized by responses to single modulation. The reduction levels were obtained for each modulation frequency (4, 5, 18, 19, 21 and 22 Hz) and each bandwidth (tone vs. noise). A 6×2 two way ANOVA was applied on the factors of modulation frequency and bandwidth, revealing that interaction between bandwidth and modulation frequency was not significant (F(5, 75) = 0.26, $p > 0.9$). Therefore neural responses to narrow and broad bandwidths were combined in the analysis of interaction index. We find that the reduction is not uniform at various modulation rates. Figure 2.3a shows that the reduction is more noticeable (~30% vs. ~15%) at fast modulations (18, 19, 21, 22 Hz) relative to slow modulations (4, 5 Hz) (significant, permutation test, $p = 0.023$). These results, taken together with the greater sensitivity to slow modulations as revealed by humans behavioral MTF, suggest an asymmetric encoding way across modulation rates.

2.3.2 Neural responses at interaction frequencies

Neural responses at interaction frequencies were also assessed by interaction level, defined as difference in neural responses by associated stimuli relative to responses by un-associated stimuli. The interaction level was obtained at each kind of interaction frequencies (difference, combination1, combination2), each bandwidth and 2 compound modulation condition (near vs. distant). A $3 \times 2 \times 2$ three way ANOVA was applied on the three factors, revealing that bandwidth did not interact with interaction frequencies and modulation condition (F(2, 30) = 1.59, $p > 0.2$). Therefore neural responses to narrow and broad bandwidth were combined in the analysis of interaction level.

We observe significant interaction activities generated by near but not by distant compound modulated stimuli (Figure 2.3b). The extent of interaction, characterized by interaction level, are comparable at all kinds of tested interaction frequencies (t-test; $p < 0.001$; $p < 0.005$; p <0.001 for three kinds of interaction frequencies) in near AM-AM condition. This differential activation between near and distant AM-AM conditions demonstrates rate separation as a critical factor in generating cortical neural responses to compound modulations, suggesting the non-linearity nature of modulation filters in the auditory cortex.

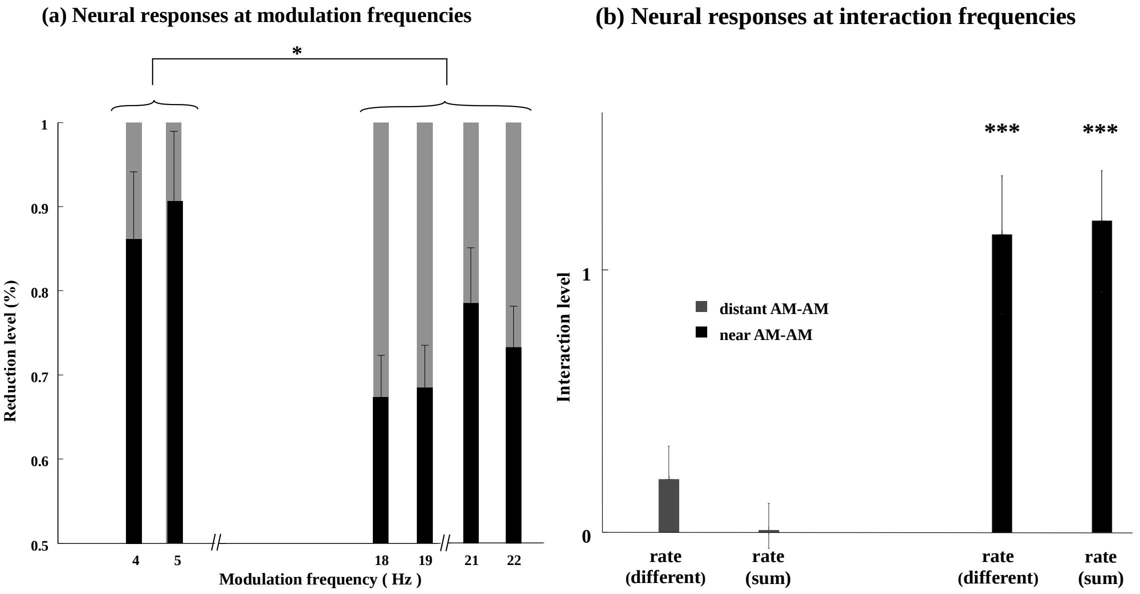

Figure 2.3: Neural response change. a) Neural response to single (respectively, compound) modulation is shown by gary (respectively, black). The neural response is reduced for sounds with compound modulations. The reduction effect is more severe at higher rates. The star indicates a significant difference between low and high rates. b) Power at interaction components. Near AM-AM sounds evoked strong interaction responses for associated sounds, relative with non-associated sounds, while distant AM-AM sounds do not evoked interaction responses.

2.4 Discussion

Neural responses to single modulated stimuli have been well established by varying various parameters, such as carrier frequency, intensity and modulation rates. The present study investigated neural responses to single and compound modulated sounds. Our results indicate that adding a second modulation to single modulated sounds causes pronounced changes in the neural response pattern which depend on the rate separation between modulations. Interaction neural activity is generated by near, rather than distant, modulation rhythms, suggesting band-limited modulation filter banks operating in the central stage of the auditory system. Furthermore, the slowest detectable neural oscillation in the auditory cortex corresponds to the perceived oscillation of the auditory percept.

2.4.1 Relations to other studies

We found that adding a second modulation reduces the response of the first modulation by 15-30%. The reduction is more prominent at fast modulations. (Dolphin, 1996, 1997) has examined responses to compound modulations in dolphins and gerbils. Focusing on higher range of modulation rates (38-142 Hz), he found a reduction of the neural response by ~30%. The maximum reductions were at the lower modulation rates (~38 Hz). The discrepancy between these results may be related to different range of modulation rates used in studies. Fast modulations (> 50 Hz) is encoded at sub-cortical levels of the auditory system, whereas slow modulation is encoded at the cortical level (Joris, Schreiner, & Rees, 2004). Essentially, Dolphin may have compared reduction effect across hierarchical levels along the auditory neural-axis, while we are focus on reduction within the auditory cortex. Our results, taken together with the greater sensitivity to slow modulations as revealed by humans behavioral MTF, suggest that cortical representations of slow modulation are more robust than fast modulation.

So far, only few studies have investigated cortical neural responses at interaction components. (Lins & Picton, 1995) found weak interaction components using 1000 Hz tone modulated by 81 Hz and 97 Hz concurrently. (John, Lins, Boucher, & Picton, 1998) employed dual modulations rates ranging from 70 to 110 Hz on different carriers and found significant interactions provided that carrier frequencies are separated by an octave and the intensities are at least 60 dB SPL. (Draganova, Ross, Borgmann, & Pantev, 2002) investigated neural responses to tones modulated by 38 & 40 Hz concurrently and found a 2 Hz component in MEG response.

The present study focuses on low modulation rates (< 25 Hz) which have been suggested to be encoded temporally in the auditory cortex. We found that stimuli modulated by slightly difference modulation rates elicited interaction components, consistent with pervious findings. However, the interaction components are absent when presented by stimuli modulated with distant modulation rates. These results clearly demonstrate rate separation as a critical factor in forming cortical representations of compound modulations. The possible neural mechanisms and physiological models are discussed in the next section.

2.4.2 Physiological models

Two alterative modulation filter models have been proposed to represent the auditory process of temporal modulation. Viemeister (1979) proposed that the demodulation of input signals is induced by the half-wave rectification and compressive processes occurring in the periphery channels. A low-pass modulation filter in subsequent auditory stages is also incorporated to account for the low-pass shape of human psychoacoustic modulation transfer function. An alterative theory postulate a modulation filter banks operating at the central stage. Distinct groups of cortical neurons tune to different modulation rates, forming modulation filter banks with different best modulation rate (Dau, Kollmeier, & Kohlrausch, 1997). In the present study, different modulations were applied to the same carrier, resulting in similar neural activation patterns in the periphery stage. Thus, the cortical neural responses to individual modulations and interaction activities suggest a function role of the central auditory process on temporal information, consistent with the latter model.

Moreover, our study estimates the bandwidths of modulation filters in this putative modulation filter bank by examining interaction responses. Sounds with distant modulation rates may be processed by distinct modulation filters, each of which has preferred/center rate. Therefore, neural responses to each modulation rate are observed (Figure 2.4a). Conversely, sounds with nearby modulation rates may activate a single modulation filter whose center rate is between the two nearby rates. In addition, due to non-linearity of the modulation filter interaction components are created in addition to the two modulation rates at the output of this filter (Figure 2.4b). Thus, the bandwidth of the modulation filter around 20 Hz should be at least 3 Hz, suggesting by the observed strong interactions evoked by compound (18 & 21Hz) modulations.

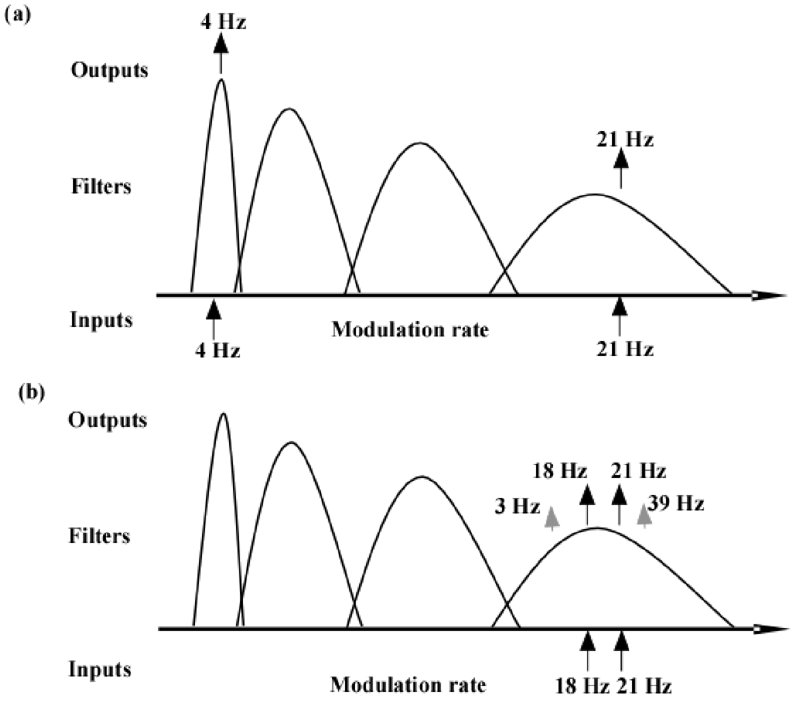

Figure 2.4: Schematic representations of modulation filter banks. The bandwidth of each filter is increased with increasing center frequency. a) Distant AM-AM is resolved by two distinct filter banks. b) Near AM-AM activate the same modulation filter bank, generating interaction components.

2.4.3 Relations to psychoacoustics

The rate of the slowest detectable neural oscillation in auditory cortex, whether associated directly with modulation of stimuli, or resulting from interactions between multiple modulations, also corresponds to the *perceived* rate of oscillation of physical stimuli. For example, sounds with both 18 & 21 Hz modulations are perceived to have a 3 Hz rhythm. Sound with both 4 & 21 Hz modulations are perceived to have a 4 Hz rhythm. The perceived oscillations of sounds with compound modulations are determined by the rate separation between modulations. Slightly different modulations produce a sensation of fluctuation at the difference rate, while distant rates produce the percept of a rhythm at the lower rate of the compound modulations. Thus, the auditory cortex tracks the slowly changing envelop of external stimuli, demonstrating its ability of integration over long time scales, complimentary to the well-known phase-locking to the fine structure.

Resolving distant modulation rates in the auditory system is of potentially importance for speech perception. It has been proposed that speech signal is segmented at two time-scales, known as syllabic rate of ~4 Hz and phoneme rate of ~20 Hz. The auditory cortex functions as a modulation filter bank with limited band-width to analyze ongoing speech on the two temporal scales simultaneously. Critically, due to the large separation between the two rates, distinct modulation filters are involved in the encoding of temporal information. In other words, no interactions are generated. Therefore the syllabic and phonetic processes work independently, not interfering with each other, allowing speech to be perceived in an unambiguous way.

Chapter 3: Informational masking study

3.1 Introduction

Attention is the cognitive process underlying our ability to focus on specific components of the environment while ignoring others. By its very definition, attention plays a key role in defining what *foreground* is, i.e. an object of attention, and differentiating it from irrelevant unattended clutter or *background* (Desimone & Duncan, 1995; J. B. Fritz, Elhilali, David, & Shamma, 2007; Gilbert & Sigman, 2007; Knudsen, 2007; Posner & Petersen, 1990). In the visual modality, studies have shown that figure/ground segmentation is mediated by a competition for neural resources between objects in the scene (Desimone & Duncan, 1995). This competition is biased in favor of different objects via top-down attention as well as behavioral and contextual effects which work to complement or counteract automatic bottom-up processes. An intricate neural circuitry has been postulated to take place in this process spanning primary visual, extrastriate, temporal and frontal cortical areas (Buschman & Miller, 2007; Ghose & Maunsell, 2002; Li, Piech, & Gilbert, 2004; Martinez-Trujillo & Treue, 2002; Qiu, Sugihara, & von der Heydt, 2007; Reynolds, Chelazzi, & Desimone, 1999; Roelfsema, Tolboom, & Khayat, 2007; Saalmann, Pigarev, & Vidyasagar, 2007).

In the auditory modality, however, there has been a limited number of studies that attempted to explore the neural underpinnings of attention in the context of auditory stream segregation, and the mechanisms governing the extraction of target sounds from a background of distracters (Hillyard, Hink, Schwent, & Picton, 1973; Hubel, Henson, Rupert, & Galambos, 1959; Tiitinen et al., 1993); Woldorff, 1993 #111; Carlyon, 2001 #47; Sussman, 2005 #109; Brechmann, 2005 #65}. It is largely unknown how top-down (e.g. task-driven or context-dependent) and bottom-up (e.g. acoustic saliency or 'pop-out') attentional processes interact to parse a complex auditory scene(Bregman, 1990; Carlyon, 2004).

In a simultaneous behavioral and neurophysiological study using magneto-encephalography (MEG), we illuminate this interaction using stimuli shown in Figure 3.1a, consisting of a repeating target note in the midst of random interferers ("maskers"). This design generalizes paradigms commonly used in informational masking experiments, (Kidd Jr, Mason, Deliwala, Woods, & Colburn, 1994), which explore how listeners' ability to perceive an otherwise salient auditory element is strongly affected by the presence of competing elements. For these stimuli, the detectability of the target note depends on various parameters, including the width of the protection zone (the spectral separation between target and masker frequencies). We adapt classic informational masking stimuli to the purposes of this study by randomly desynchronizing all background maskers throughout the duration of the stimulus, making the target note the only regular frequency channel in the sequence (with repetition rhythm of 4 Hz). While maintaining the same physical stimulus, we contrast the performance of human listeners in two complementary tasks: **(i)** a "target task", in which subjects are asked to detect a frequency shifted (ΔF) deviant in the repeating target signal; and **(ii)** a "masker task", in which subjects are asked to detect a sudden temporal elongation (ΔT) of the masker notes.

Crucially, attention is required to perform either task, but the subjects' attention must be focused on different features of the acoustic stimuli each case.

3.2 Methods

3.2.1 Subjects

Nine subjects (6 males; mean age 29 years) participated in the psychoacoustic study. Eighteen subjects (11 males; mean age 27 years) participated in the MEG study. Three subjects took part in both studies. Among the eighteen subjects in the MEG study, four subjects were excluded from further analysis due to an excess of non-neural electrical artifacts or an inability to perform the tasks. All subjects were right handed (Oldfield, 1971), had normal hearing and no history of neurological disorder. The experiments were approved by the University of Maryland Institutional Review Board and written informed consent was obtained from each participant. Subjects were paid for their participation.

3.2.2 Stimulus design

The stimuli were generated using MATLAB (MathWorks Inc, Natick MA). Sounds were 5.5 s in duration and sampled at 8 kHz. Each stimulus contained one target note, repeating at 4 Hz, whose frequency was randomly chosen in the range 250-500 Hz in two semitone intervals. The background consisted of random tones at a density of 50 tones/s, uniformly distributed over time and log-frequency (except for the spectral protection zone). The frequencies of the random notes were randomly chosen from the five octave range centered at 353 Hz, in two semitone intervals, with the constraint that no masker components were permitted within a 4-, or 8-, or 12-semitone around the target frequency (the protection zone half-width). Each masker pair had a minimum spectral spacing of 2 semitones. Masker and target tones were 75 ms in duration with a 10 ms onset and offset cosine ramps.

Fifteen exemplar stimuli were generated for each of the four condition types: null condition (no deviants); target condition (one target deviant per stimulus); masker condition (one masker deviant per stimulus); and combined condition (one target deviant and one masker deviant, at independent times, per stimulus). Each target deviant was the displacement of a target note (upward or downward) by 2 semitones from the target frequency. Each masker deviant was a single 500 ms time-window in which all masker tones were elongated from 75 ms to 400 ms. The temporal location of the deviant (for both target and masker tasks) was randomly distributed along the 5.5 s stimulus duration, with timing as indicated by the behavioral buildup curve in Figure 3.5.

3.2.3 Experimental procedure

In the psychoacoustic experiment, participants were presented with 180 stimuli (3 protection zones × 4 conditions × 15 exemplars) per block. The progression from one stimulus to the next was initiated by the subject with a button-press. The stimulus set in the MEG experiment was identical to the psychoacoustic stimuli, except that only one (8 semitone half-width) protection zone was used, giving 60 stimuli (1 protection zone × 4 conditions × 15 exemplars) per block. The inter-stimulus interval (ISIs) were randomly chosen to be 1800, 1900, or 2000 ms. Three blocks were presented, and between blocks subjects were allowed to rest but were required to stay still.

The identical stimulus ensemble (including identical ISIs in the MEG case) was presented for both target and masker tasks. Depending on the task being performed, subjects were instructed to listen for the presence of a frequency deviant in the target rhythm (target task) or a duration deviant in the masker (masker task); each task deviant was present in exactly half the stimuli.

3.2.4 Psychoacoustical study

Subjects were seated at a computer in a soundproof room. The signals were created off-line and presented dichotically through Sony MDR-V700 headphones. Subjects controlled the computer using a Graphical User Interface (GUI) using the mouse. The task was described to subjects as well as the basic use of the GUI. Subjects were allowed to adjust the volume to a comfortable level before proceeding with the experiment.

Each subject performed both the masker task and the target task, with task order counterbalanced across subjects. Each task required the subject to listen to one block of the 180 stimuli described above. Each stimulus was presented only once and no feedback was given after each trial. The entire session of both tasks lasted approximately 1.5 hours.

A training block of 20 stimuli was presented before each task block. In the target task training, the protection zone half-width decreased from 12 semitones to 4 semitones in steps of 4 semitones. In the masker task training, it increased from 4 semitones to 12 semitones in steps of 4 semitones. Subjects were permitted to listen to each sound as many times as desired; then participants were prompted to indicate whether a deviant was present. The correct answer was displayed afterwards. Subjects pressed a button to initiate the presentation of the next stimulus.

3.2.5 MEG study

Subjects were placed horizontally in a dimly lit magnetically shielded room (Yokogawa Electric Corporation, Tokyo, Japan). Stimuli were presented using Presentation software (Neurobehavioral Systems, Albany, CA). The signals were delivered to the subjects' ears with 50 Ω sound tubing (E-A-RTONE 3A, Etymotic Research, Inc), attached to E-A-RLINK foam plugs inserted into the ear-canal, and presented at a comfortable loudness of approximately 70 dB SPL.

Before the main experiment, a pre-experiment was run, where a 1 kHz, 50 ms tone pip was presented about 200 times. The ISI was randomized between 750 ms and 1550 ms and subjects were instructed to count the tone pips. The aim of this task was to record the M100 response (a prominent peak approximately 100 ms after pip onset, also called N1m) used for differential source localization. The responses were checked to verify that the location and strength of neural signals fell in a normal range.

In the main experiment, participants were presented with three blocks of the 60 stimuli described above. Each subject performed both the masker task and the target task, with task order counterbalanced across subjects. Subjects were instructed to press a button held in the right hand as soon as they heard the appropriate deviant.

A training block with 12 sounds was presented before each task. Each training sound was presented twice. Subjects verbally indicated the existence of the deviants and the feedback was given by the investigator. The entire session of both tasks lasted approximately 1 hour.

MEG recordings were conducted using a 160-channel whole-head system (Kanazawa Institute of Technology, Kanazawa, Japan). Its detection coils are arranged in a uniform array on a helmet-shaped surface of the bottom of the dewar, with about 25 mm between the centers of two adjacent 15.5 mm diameter coils. Sensors are configured as first order axial gradiometers with a baseline of 50 mm; their field sensitivities are 5 fT/√Hz or better in the white noise region. Three of the 160 channels are magnetometers separated from the others and used as reference channels in noise filtering methods. The magnetic signals were bandpassed between 1 Hz and 200 Hz, notch filtered at 60 Hz, and sampled at the rate of f_s = 1000 Hz. All neural channels were denoised twice with a Block Least Mean Square (LMS) adaptive filter: first using the 3 external reference channels (Ahmar & Simon, 2005), and secondly using the 2 channels with the strongest cardiac artifacts (Xiang, Wang, & Simon, 2005).

3.2.6 Data analysis

Behavioral Performance Analysis

The ability of subjects to perform the requested task was assessed by calculating a d-prime measure of performance (Kay, 1993). For each condition (i.e. each task and protection zone), we estimated the correct detection and false alarm probabilities for detecting the target or masker deviants; converted them to normal deviates (z-scores) and computed the d-prime value. The performance shown in Figure 3.1b depicts the mean d-prime values across subjects. The error bars represent the standard error of mean.

To determine the effect of the target's tonal frequency on the neural responses, the stimuli were divided spectrally: each sound was characterized as a low or high frequency target tone sequence depending on the target tone's relation to the middle frequency 353 Hz (those with target tone frequency 353 Hz were randomly assigned as low or high in such a way to equipartition the high and low categories). A d-prime measure was then derived for each of the low or high target trials from both target and masker tasks.

To investigate the buildup of the target object during the target task, we divided the deviant trials according to the 9 possible temporal locations of the deviant throughout the stimulus sequence. A probability of hit was then measured for each trial. Because of the temporal uncertainty in the false alarm trials, we calculated an average false alarm rate (irrespective of when the false response was issued), and combined it with the time-specific hit rate to derive a d-prime measure for each time segment. Using this behavioral assessment measure, five subjects yielded non-positive d-prime values due to their high false alarm rate and low hit rate and were excluded from the analysis of buildup.

Neural Data Analysis

After recordings were completed and noise reduction algorithms applied, the analysis epochs of each task were extracted from 1.25 seconds to 5.5 seconds post-stimulus and concatenated, forming a single response with duration T = 765 seconds

(4.25 seconds × 60 sounds × 3 blocks) in duration. The Discrete Fourier Transform (DFT) was applied on the single response, giving a single Fourier response of from 0 to 500 Hz with frequency resolution 1/765 Hz.

The evoked neural responses to the target sequences were characterized by the magnitude and phase of the frequency component at 4 Hz (the tone presentation rate) and were used for localization and for phasor maps. The complex magnetic field strength is given by the product of the value of DFT times the sampling interval ($1/f_s$), and has units of fT/Hz. Power spectral density is calculated as the product of the inverse duration ($1/T$) times the modulus squared of the complex magnetic field strength, and has units of fT^2/Hz. The remainder of the analysis was based on the normalized neural responses, defined to be the squared magnitude of the frequency component at 4 Hz divided by the average squared magnitude of the frequency components between 3 Hz and 5 Hz (excluding the component at 4 Hz), averaged over the 20 channels with the strongest normalized neural responses. Using 10, 20 or 50 channels yielded similar findings; however, only the 20 channel analysis is reported here.

The spatial pattern of the neural responses was represented by a phasor map, a graph of the complex (magnitude and phase) magnetic field on all channels. For each channel, the length of the vector arrow is proportional to the magnitude of 4 Hz frequency component and the direction of the arrow represents the phase according to standard polar coordinates. Red and green contours represent the magnetic field strength projected onto the line of constant phase that maximizes the projected field's variance

(Simon & Wang, 2005). The phasors are visually faded using the signal-to-noise ratio (SNR) of each channel as linear fading coefficients.

The normalized neural responses difference between target task and masker task was averaged across 14 subjects to characterize attention gain effect. Furthermore, to evaluate the effect of attention at across frequencies, the same analysis is done at 4 Hz and the two adjacent frequency bins (4 Hz - Δ*f*, 4 Hz + Δ*f*), and also at 11 frequencies in the alpha, theta and low gamma frequency bands, in ~5 Hz increments from ~5 Hz to ~55 Hz. For consistency with the frequencies examined in tests of phase coherence (next), we used $\Delta f = 1/(4.25)$ Hz, with analysis performed only at its integer multiples (e.g. $17\Delta f = 4.0$ Hz and $21\Delta f \approx 4.94$ Hz).

To study attention modulation effects on the synchronization between two distinct neural populations, phase coherence between channels *m* and n, $\gamma^2{}_{mn}$ is obtained from Q=180 trials (Bendat & Piersol, 1986; Srinivasan, Russell, Edelman, & Tononi, 1999)

$$\gamma^2{}_{mn}(f) = \frac{\left|X_{mn}(f)^2\right|}{\left\langle X_{mm}(f)\right\rangle\left\langle X_{nn}(f)\right\rangle}$$

where, $X_{mn}(f)$ is average cross spectrum between channel *m* and channel n, $X_{mm}(f)$ is average power spectrum of the individual channel *m*:

$$X_{mn}(f) = \frac{1}{Q}\sum_{q=1}^{Q} F_{mq}(f)F_{nq}(f)^*$$

where $F_{mq}(f)$ is the Fourier transform of the qth trial of channel m at frequency f. A coherence value of 1 indicates that the two channels maintain the same phase difference on every trial, whereas a coherence value near 0 indicates a random phase difference across trials. The coherence difference between target task and masker task was computed for every channel pair. The SEM ε_{mn} was constructed to identify robust coherence change (Bendat & Piersol, 1986; Srinivasan, Russell, Edelman, & Tononi, 1999).

$$\varepsilon_{mn} = \sqrt{\frac{2}{Q}}\left(\frac{1-\gamma^{2}{}_{mn}}{\left|\gamma_{mn}\right|}\right).$$

To emphasis phase modulation in auditory cortex, each subject's 20 channels with the strongest normalized neural response at target rate are included in further analysis. In addition, to exclude the artificial coherence resulted from volume conduction effects on extracranial magnetic field and measure genuine phase correlation between distinct populations of neurons, only long distance channel pairs (channel separation > 100 mm) were included (Srinivasan, Russell, Edelman, & Tononi, 1999). The difference between number of channel pairs with robust increased coherence and channel pairs with decreased coherence is normalized over the total number of long range channel pairs for each subject. Furthermore, to evaluate the effect of attention at across frequencies, the same analysis is done at 4 Hz and the two adjacent frequency bins (4 Hz - Δf, 4 Hz + Δf), and also at 11 frequencies in the alpha, theta and low gamma frequency bands, in ~5 Hz increments from ~5 Hz to ~55 Hz. For phase coherence (measured across trials), Δf =

1/(4.25) Hz, and phase coherence was analyzed only at its integer multiples (e.g. $17\Delta f = 4.0$ Hz and $21\Delta f \approx 4.94$ Hz).

To investigate the possibility of hemispheric bias, the 20 channels with the strongest normalized neural response at the target rate were chosen from left and right hemisphere respectively to represent the overall neural activity of each hemisphere. Neural responses averaged across the 20 channels are subtracted across hemispheres for each task and for all 14 subjects. Using 10, 20 up to 60 channels yielded similar findings; however, only the 20 channel analysis is reported here

To determine the effect of the tonal frequency on the neural responses, the stimuli were divided spectrally as described above. The neural responses at the target rate (both normalized response and phase coherence), from low and high frequency target tone stimuli, were obtained for each subject in the same way as described above, but with only appropriate epochs concatenated or phase-averaged.

To investigate the buildup of the target object in target task, the responses were divided temporally: the analysis epochs were divided into five temporal segments with 750 ms duration each and corresponding segments were concatenated across epochs. The first segment began at 1250 ms post stimulus since earlier time intervals showed substantial power at the frequency corresponding to the segment duration, an artifact indicating that the measured spectral power was extrinsic to the analysis window, not intrinsic to the neural signal. The segment duration 750 ms was used since shorter durations did not show the buildup effect, indicating that the neural response reflects phase-locked activity entrained at the target rate. An analogous analysis of phase coherence over time was performed but did not yield significant results.

Behavioral vs. Neural Correlation and Bootstrap Analysis

We correlated the effect of high vs. low target frequencies in the behavioral and neural responses by contrasting the per-subject psychometric and neurometric measures. First, we scaled the neural data (i.e. the normalized responses to target) by a factor of 3 in order to match the absolute ranges of both neural and behavioral values. We then derived the angle (i.e. inverse tangent) of the slope relating the high vs. low frequencies of the behavioral and neural data points for each subject and each task. The across subject slopes were then combined using circular statistics to yield an angular mean for each task (Fisher, 1993).

We performed a bootstrap procedure in order to confirm the positive (respectively, negative) correlation between the neurometric and psychometric functions in the target (respectively, masker) task. We followed a balanced bootstrap sampling procedure (Efron & Tibshirani, 1993) by randomly selecting 14 subjects with replacement and computing their angular sample mean and repeating this process 1000 times. The procedure was controlled to ensure that all subjects appear the same number of times over all 1000 bootstrap samplings. Confidence measures were then derived from the bootstrap statistics.

A similar statistical analysis was performed to correlate the psychometric and neurometric curves for the target detection buildup. To match the range of values from the neural and behavioral data, we scaled the neural responses by a factor of 2 (note that the different scaling is due to the reduced values of the normalized neural response due to the smaller window for the buildup analysis). The behavioral curves for each subject were then interpolated to match the sampling rate of the neural data. Subsequently, these two curves were then fitted by a first-order polynomial to derive the slope relating the two functions. The slope value was transformed into an angle, and combined across subjects following the same procedure described above. Note that the five subjects with negative d-prime values were excluded from this correlation analysis, because of their questionable behavioral performance.

Neural Source Localization

Source localization for the M100 response was obtained by calculating the current-equivalent dipole best fitting the magnetic field configuration at the M100 peak, in each hemisphere. Source localization for the neural response to the target was obtained by calculating the complex current-equivalent dipole best fitting the complex magnetic field configuration at 4 Hz peak, in each hemisphere (Simon & Wang, 2005). Only channels with SNR > 4 were used in the fitting. Significance of the relative displacement between the M100 and 4 Hz dipole sources were determined by a two-tailed paired t-test in each of three dimensions: lateral/medial, anterior/posterior, and superior/inferior.

3.3 Results

The effect of protection zone width on the performance of both tasks is illustrated in Figure 3.1b. In the left panel it can be seen that the detectability of the target becomes easier with increasing protection zone (significantly positive slope; bootstrap across subjects, $p < 10^{-4}$), a result that is in line with previous hypotheses of streaming that correlate the ease of target detection with the frequency selectivity of neurons in the central auditory system (Fishman, Reser, Arezzo, & Steinschneider, 2001; Fishman & Steinschneider, 2006; Kowalski, Depireux, & Shamma, 1996; Micheyl, Tian, Carlyon, & Rauschecker, 2005; Miller, Escabi, Read, & Schreiner, 2002).

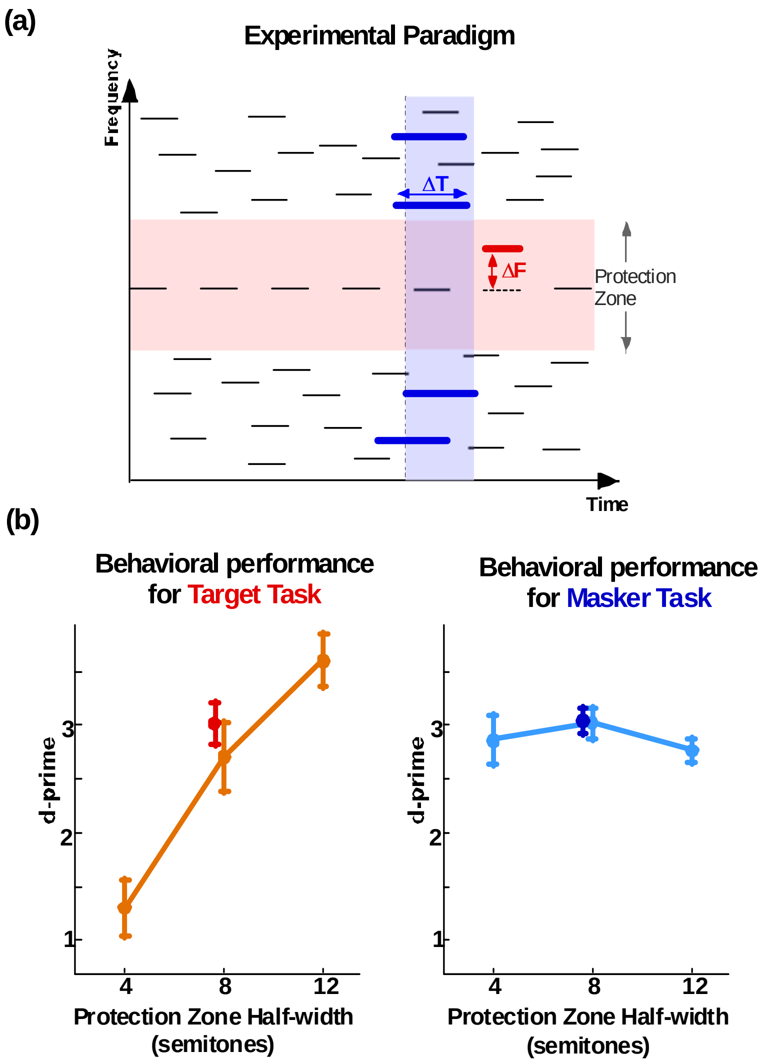

Figure 3.1: Stimulus Description and Behavioral Performance. a) Cartoon spectrogram of a typical stimulus. The stimulus consists of a repeating target note embedded in random interferers. A protection zone surrounds the target frequency with a spectral width of twice the minimal distance between the target note

and nearest masker component (orange band). In the target task subjects are instructed to detect a frequency shifted (ΔF) deviant in the repeating target notes. In the masker task subjects are instructed to detect a sudden temporal elongation (ΔT) of the masker notes. b) Behavioral performance results for target and masker tasks, as measured by d-prime as a function of protection zone width. Orange (respectively, light blue) lines show the mean performance in task-detection in the target task (respectively, masker task) in the psychoacoustical study. Red (respectively, dark blue) points show the mean performance in task detection in the target task (respectively, masker task) in the MEG study (8 semitone condition only). Error bars represent standard error.

In contrast, the same manipulations of protection zone do not substantively affect masker task performance (right panel) (not significantly different from zero; bootstrap across subjects $p > 0.3$). The masker task, designed to divert attentional resources away from the target, involves a more diffuse attention to the spectrally broad and distributed masker configuration, and, compared to the target task, reflects a different top-down bias in the way the same stimulus is parsed. The behavioral performance was unchanged whether tested under purely psychoacoustic or neural recording conditions (no significant difference; unpaired t-test; target task: $t = -0.75$, $p = 0.46$; masker task: $t = 0.09$, $p = 0.93$).

Depending on listeners' attentional focus, the percept of an auditory target in a complex scene is *differentially* mirrored by the responses of neurons in auditory cortex. Using the high temporal resolution of MEG, we measure the neural responses to this stimulus paradigm in 14 human subjects. Figure 3.2a reveals that, during the performance of the target task, the target rhythm emerges as a strong 4 Hz component in the neural signal of an individual subject. In contrast, during the masker task the cortical response entrained at 4 Hz is noticeably suppressed in comparison (Figure 3.2a, right panel). This differential activation is strong evidence of the modulatory effect of task-dependent attention on the neural representation of a single acoustic stimulus, much like visual attention (Maunsell & Treue, 2006; Reynolds & Chelazzi, 2004). Additionally, this attentional effect on the neural signal is not just momentary but is *sustained* over the duration of the stimulus sequence.

This attentional effect is confirmed in the population of 14 subjects (Figure 3.2b), with an average normalized neural response of 20.9 in the target task and 8.3 in the masker task: a gain of more than two and a half for neural phase-locked, sustained activity when subjects' attention is directed towards the repeating note (individually, 11 out 14 subjects show a significant increase: unpaired t-test, $p < 10^{-4}$). Direct correlation between the target task neural response and target task behavior is not observed, but as shown below, *changes* in a subject's target neural response are significantly correlated with *changes* in the subject's behavioral responses.

The MEG magnetic field distributions of the target rhythm response component, examples of which are shown at the inset of graphs in Figure 3.2a, reveal the stereotypical pattern for neural activity originating separately in left and right auditory cortex. The neural sources of all the target rhythm response components with sufficiently high signal-to-noise ratio originate in auditory cortex (Simon & Wang, 2005). The neural source's mean displacement from the source of the auditory M100 response (Naatanen & Picton, 1987) was significantly different (two-tailed t-test; $t = 2.9$, $p = 0.017$) by 14 ± 5 mm in the anterior direction, for the left auditory cortex only (no significant differences were found in the right hemisphere due to higher variability there). Assuming an M100 origin of *planum temporale*, this is consistent with an origin for the neural response to the target rhythm in Heschl's gyrus, the site of core auditory cortex.

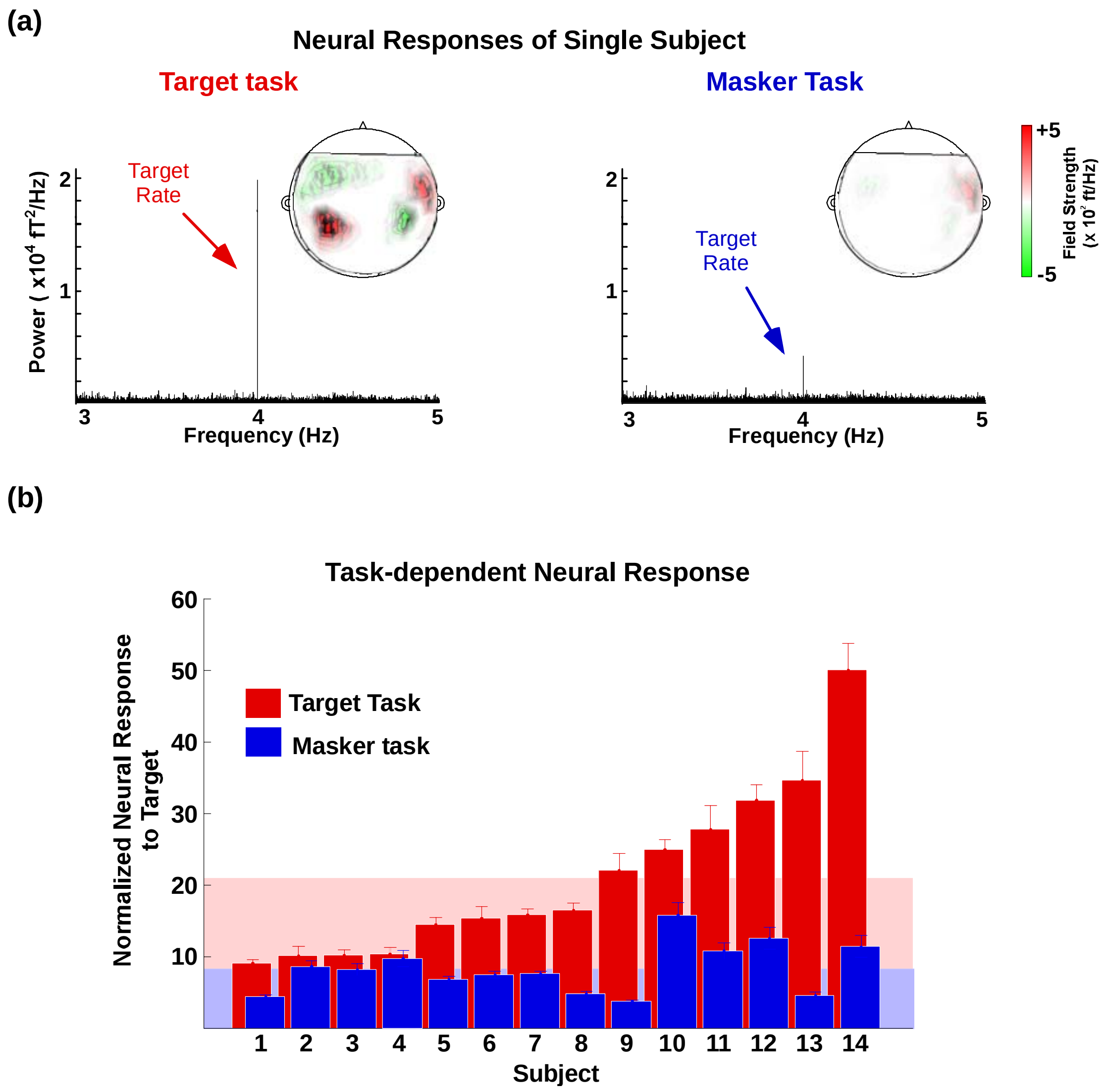

Figure 3.2: Neural Responses. a) Power spectral density of MEG responses for a single subject (subject 14 in Figure 2b below) in target (left) and masker (right) tasks, averaged over twenty channels. Insets: The MEG magnetic field distributions of the target rhythm response component. Red and green contours represent the target magnetic field strength projected onto a line with constant phase. b) Normalized neural response to the target rhythm by subject (individual bars) and task (red for target task, blue for masker task). The normalized neural response is computed as the ratio of the neural response power at the target rate (4

Hz) to the average power of the background neural activity (from 3–5 Hz; see METHODS). Bar height is the mean of the twenty best channels; error bars represent standard error. Light pink background (respectively, light blue) is the mean over subjects for the target task (respectively, masker task).

The neural response change at the target rate of 4 Hz is highly significant (bootstrap across subjects, $p < 10^{-4}$) (Figure 3.3a). In contrast, there is no significant change in normalized neural response at other frequencies, whether at frequencies nearby (one frequency bin on either side of 4 Hz) or distant (alpha, theta, and low gamma band frequencies sampled with approximately 5 Hz spacing up to 55 Hz). This demonstrates that this feature-selective sustained attention modulates the cortical representation of the specific feature, but not general intrinsic rhythms, whether in the same band or other bands.

Changes in response phase coherence across channels was also assessed at the same frequencies (Figure 3.3b, sample subject in Figure 3.3c). Only the phase coherence at the target rate shows a significant enhancement (bootstrap across subjects, $p = 0.002$), further demonstrating that change from one form of attention to another does not modulate general intrinsic rhythms. This 30% enhancement is distributed across channel pairs, revealing increased phase coherence both within and across hemispheres.

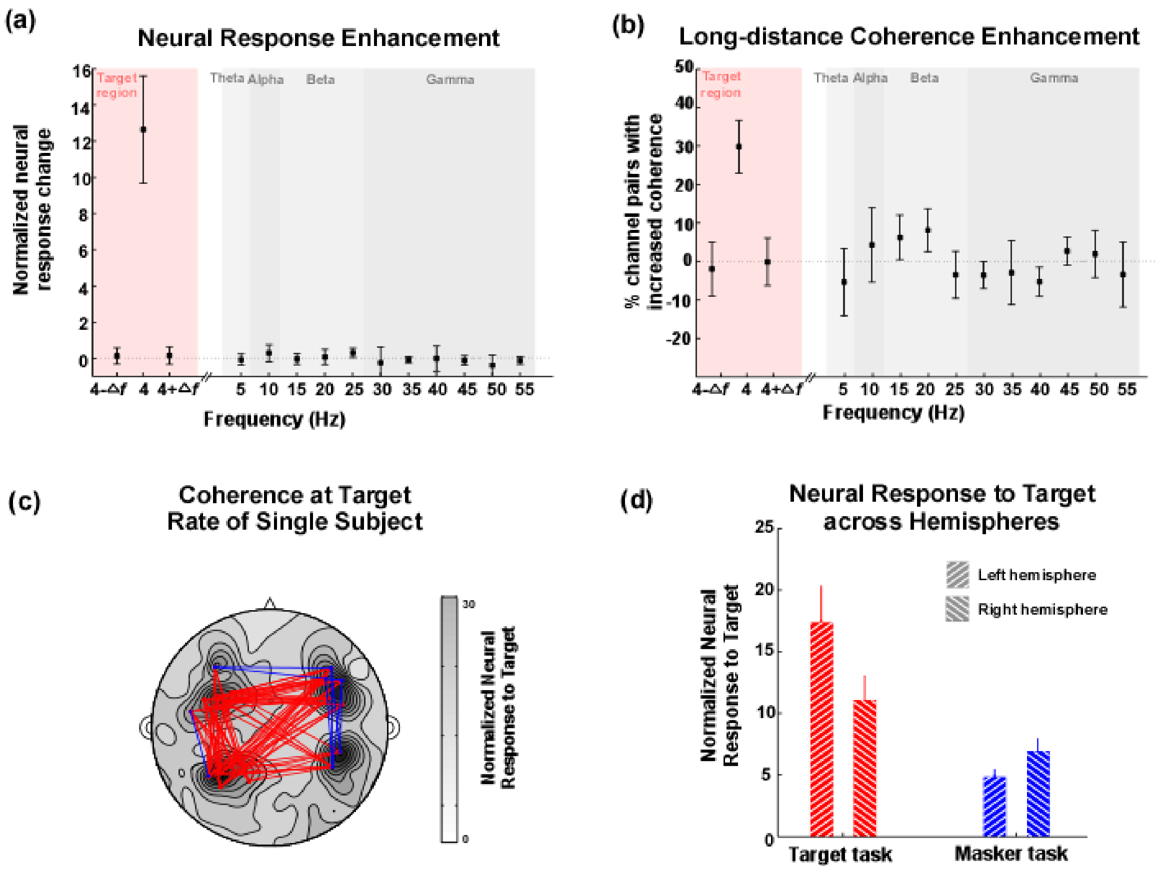

Figure 3.3: Power and phase enhancement during target task. a) Normalized neural response of target task relative to masker task shows enhancement exclusively at 4 Hz (the frequency of the target rhythm). Each data points represents the difference between normalized neural response of target relative to masker task, error bars represent standard error The asterisk at 4 Hz shows that only that particular frequency yields a statistically significant enhancement. b) Phase coherence between distant MEG channels of target relative to masker task. The difference between number of long range channel pairs with robust increased coherence in target task and channel pairs with decreased coherence is normalized over the total number of long range channel pairs. The phase enhancement is significant (shown with asterisk) only at 4 Hz. c) Channel pairs with robust coherence difference at target rate for single subject, overlayed on the contour map of normalized neural response at target rate. Each channel pair with enhancement coherence is connected by a red line, while pairs with decreased coherence are connected by a blue line. Coherence is only analyzed for the twenty channels with best normalized response to target rhythm. d) Neural responses to target across hemispheres. The 20 channels with the strongest normalized neural response at target rate

are chosen from left and right hemisphere respectively to represent the overall neural activity of each hemisphere. Neural responses averaged across the 20 channels and 14 subjects are compared across hemispheres and tasks. The left hemisphere shows stronger activation at target rate in target task, while right hemisphere shows stronger activation in masker task.

We also observe a task-dependent hemispheric asymmetry in the representation of the neural response at the target rate. During the target task, the left hemisphere shows a greater normalized neural response than the right hemisphere (bootstrap across subjects, $p = 0.001$); during the masker task the right hemisphere shows a greater normalized neural response than the left hemisphere (bootstrap across subjects, $p = 0.04$) (Figure 3.3d).

Together with the behavioral demands of the task, the *bottom-up saliency* of a target note contributes to both the neural response and subject performance. A close examination of the physical parameters of the stimulus reveals that the frequency of the target note affects the audibility of the repeating rhythm, with higher frequency targets popping out more prominently than their lower frequency counterparts. This variation in the pop-out sensation may be explained by the contours of constant loudness of human hearing showing an approximately 5 dB decrease over the target note range 250-500 Hz (ISO-226, 2003). We exploit this physical sensitivity of the auditory system and determine the effect of this target pop-out on the neural and behavioral performances in both target and masker tasks. Figure 3.4a (orange curve) confirms that behavioral performance in the target task is easier for higher frequency targets (> 350 Hz) than for lower frequencies (t-test; $t = -3.3$, $p = 0.002$). Correlated with this trend is an increased neural response to the target for higher frequencies compared to lower frequencies (red curve) (increase not statistically significant alone). Conversely, the masker task shows a trend of being oppositely affected by the physical saliency of the target note despite its irrelevance for task performance (approaching significance; t-test, $t = 1.8$, $p = 0.08$). On the one hand, the neural power is increased for high-frequency target reflecting their increased audibility (dark blue curve) (though not statistically significant alone). On the other hand, as the target becomes more prominent, the subjects' performance of the background task deteriorates indicating a distraction effect caused by the presence of the repeating note (light blue curve). Additionally, phase coherence is significantly enhanced for high-frequency targets over low-frequency targets only during the target task (bootstrap across subjects, $p < 10^{-3}$) (Figure 3.4c). This result confirms that the physical parameters and acoustic saliency of a signal can interfere with the intended attentional spotlight of listeners and effectively deteriorate task performance (Gottlieb, Kusunoki, & Goldberg, 1998; Naatanen, Tervaniemi, Sussman, Paavilainen, & Winkler, 2001), both neurally and behaviorally.

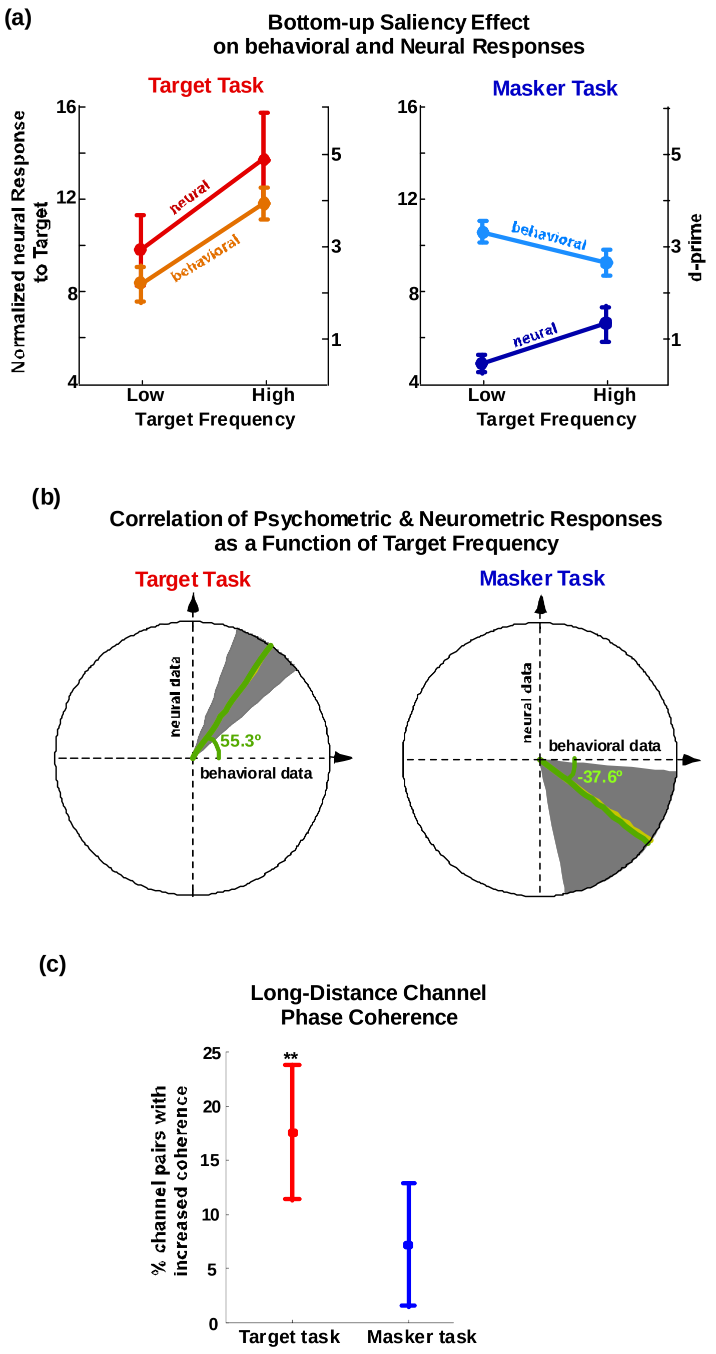

Figure 3.4: The effect of bottom-up acoustic saliency on behavior and neural responses. a) Normalized neural response to target rhythm, and behavioral performance, as a function of target frequency in target task (left) and masker task (right), averaged over subjects. Error bars represent standard error. b) Correlation of behavioral and neural responses as a function of target frequency. The ratio of the neural to behavioral response differences as a function of target frequency, interpreted as a slope angle, is averaged

across subjects yielding a mean slope angle of 55.1° for target (left) task and -36.3° for masker (right) task (yellow line). Bootstrap estimates (overlaying green lines) and their 95% confidence intervals (gray background) confirm the positive (respectively, negative) correlations for target (respectively, masker) task. c) Phase coherence between distant MEG channels of target relative to masker task for high-frequency targets over low-frequency targets. High-vs.-low frequency targets show significant enhancement only for target task.

In order to establish the correspondence between the neural and behavioral responses under both task conditions in a parametric way, we quantified the slope (converted into an angle) relating the normalized neural signal with the listener's d-prime performance on a per-subject basis. The average slope angle for the target task is 55.1º, i.e. positive slope, demonstrating the positive correlation between the two measures. Bootstrap analysis confirms this; Figure 3.4a, left panel, illustrates both the bootstrap mean of 55.3º (green line) and the 5th to 95th percentile confidence limits (gray background), all with positive slope. Analysis of the masker task also demonstrates the anti-correlation trend between the neural and behavioral data, with an average slope angle of -36.3º shown in yellow. The bootstrap analysis also confirms this; Figure 3.4b (right panel) shows that the 5th to 95th confidence intervals (gray background) yield a robust negative slope with a bootstrap mean of -37.6 º (green line).

The perceptual detectability of the regular target rhythm improves over time, following a pattern that is highly correlated with the neural buildup of the signal representation. Consistent with previous findings of buildup of auditory stream segregation (Anstis & Saida, 1985; Bregman, 1978), subjects' performance during the target task improves significantly over several seconds as shown in Figure 3.5a (solid orange line) (bootstrap across subjects, $p < 10^{-4}$). This similarity suggests that target detection is mediated by mechanisms analogous to those employed in auditory streaming and object formation (Griffiths & Warren, 2004). Moreover, the neural response to the target rhythm also displays a statistically significant buildup (Figure 3.5a, dashed red line) (bootstrap across subjects, $p = 0.02$) closely aligned with the behavioral curve, and, consequently, decoupled from the actual acoustics. The remarkable correspondence between these two measures strongly suggests that the enhanced perception of the target over time is mediated by an enhancement of the neural signal representation, itself driven by an accumulation of sensory evidence.

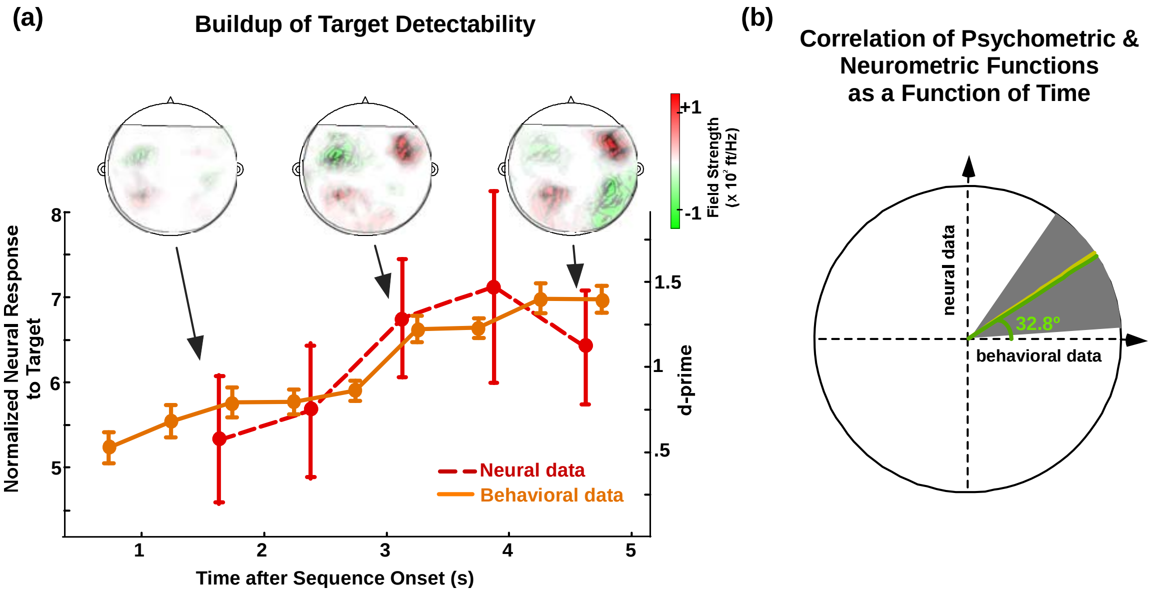

Figure 3.5: Buildup over time of behavioral and neural responses in target task. a) Normalized neural response to target rhythm, and behavioral performance, as a function of time in target task averaged over subjects. Error bars represent standard error. Insets: The MEG magnetic field distributions of the target rhythm response component for a single subject at representative moments in time (subject 10 from Figure 2b). b) Correlation of behavioral and neural responses as a function of time. The ratio of the neural to behavioral response trends as a function of time, interpreted as a slope angle, is averaged across subjects

yielding a mean slope angle of 34.3° (yellow line). Bootstrap estimates (overlaying green line) and the 95% confidence intervals (gray background) confirm the positive correlation between the psychometric and neurometric buildup curves.

The MEG magnetic field distributions of the target rhythm response component in Figure 3.5a (insets), showing the stereotypical pattern of neural activity originating separately in left and right auditory cortex, illustrate the changing strength of the neural activity over time in an individual subject.

We confirm the correlation between the psychometric and neurometric curves over time by running a bootstrap analysis on a per-subject basis. As expected, the slope correlating the d-prime and neural response curves for each subject yield a mean positive slope angle of 34.3º; bootstrap across subjects shows a mean of 32.7º, with the 5th to 95th confidence intervals falling within the upper right quadrant (Figure 3.5b). We also note that the sub-segments over which the neural buildup is measured are required to span several rhythmic periods (at least three). There is no buildup using intervals with shorter durations, despite sufficient statistical power, implicating *temporal* phase coherence (in contrast to spatial phase coherence) as critical to the buildup of the neural target representation.

3.4 Discussion

This study's novel experimental paradigm builds on previous work in stream segregation using simpler stimuli (Fishman, Reser, Arezzo, & Steinschneider, 2001; Gutschalk et al., 2005; Micheyl, Tian, Carlyon, & Rauschecker, 2005; Snyder, Alain, &

Picton, 2006) but (**i**) using a richer stimulus design and (**ii**) keeping the physical parameters of the stimulus fixed while manipulating only the attentional state of the listeners. The major finding is that auditory attention strongly modulates the sustained neural representation of the target. Specifically, sustained attention correlates with a sustained increase in the time-varying neural signal, in contrast with transient (Naatanen, 1990; Tiitinen et al., 1993) or non-specific, constant ("DC") (Hari, Aittoniemi, Jarvinen, Katila, & Varpula, 1980; Picton, Woods, & Proulx, 1978a, , 1978b) effects of attention on auditory signals. This sustained increase in the signal strength is specific to the frequency of the target rhythm, and is additionally complemented by an enhancement coherence over distant channels, reflecting an increased synchronization between distinct underlying neural populations. This attentional effect (in both power and phase) appears exclusively at the target frequency and is absent even from adjacent frequency bins; which argues against any theory of neural recruitment or redistribution of energy at the low frequency spectrum. Therefore, our findings argue that processes of attention interact with the physical parameters of the stimulus, and can act exclusively to enhance particular features to be attended to in the scene, with a resolution of a fraction of a Hz.

Secondly, the data reveals that enhanced acoustic saliency, which causes an increase in perceptual detectability, also correlates with an increase in the sustained power and coherence of the neural signal. In this case, the increase in neural signal occurs regardless of the task being performed, but with different behavioral consequences: in the target task, it leads to an increase in performance, but in the masker task, a decrease (via interference). This outcome allows us to give different explanations of this 'attentionallymodulated’ neural change: as a marker of object detectability during the first task; but as a neural correlate of perceptual interference during the second task.

Thirdly, the data show a left hemisphere bias in the cortical representation of the target, for the target task, suggesting a function role of left hemisphere in selective attention, consistent with previous findings in visual (Zani & Proverbio, 1995) and auditory (Bidet-Caulet et al., 2007; Coch, Sanders, & Neville, 2005) modalities. In contrast, for the masker task, the hemispheric bias in cortical representation of the (now non-attended) target is reversed to the right, and might be simply due to the nature of the attentional demands of the task (more diffuse attention to the global structure of the sound), the temporal nature of the deviant in the masker task (elongation of the background tone), or to the general ASSR right-hemispheric bias.

Finally, this study offers the first demonstration of the buildup over time of the neural representation of a target signal that also follows the same temporal profile of the buildup based on listeners’ detectability performance. Using the current experimental paradigm, we are able to monitor the evolution in time of attentional processes as they interact with the sensory input. Many studies overlook the temporal dynamics of the neural correlates of attention, either by using cues that prime subjects to the object of attention (thereby stabilizing attention before the onset of the stimulus), or by explicitly averaging out the build-up of the neural signal in their data analysis (focusing instead on the overall contribution of attention in different situations, and not monitoring the dynamics by which the process builds up). Our findings reveal that even though the sensory target signal is unchanged, attention allows its neural representation to grow over time, closely following the time-course of the perceptual representation of the signal.

Together, these findings support a view of a tightly coupled interaction between the lower level neural representation and the higher level cognitive representation of auditory objects, in a clear demonstration of the cocktail party effect (Cherry, 1953). Our experimental paradigm allows both task-driven (top-down) and stimulus-driven (bottom-up) processes to guide perception. For listeners performing the target task, the target rhythm is the attended auditory object, a foreground stream to be separated from a noisy background. The masker task, requiring the listener to reverse the role of the foreground and background, allows the contrasting situation to be considered under otherwise identical acoustical conditions. This permits a controlled de-emphasis of the auditory role of the target rhythm, without the need for a "passive" listening condition under which the amount of the listener's attention is lessened but actually unknown and strongly variable across subjects.

The data suggest that new models of attention may be required, based on *temporally* coherent or locally synchronous neural activity rather than neural amplification (Niebur, Hsiao, & Johnson, 2002). The buildup of neural responses over time is seen only when integrated over several periods of the target rhythm, but not for individual periods. This result is difficult to explain using standard models of attention that rely solely on gain-based changes, or even on gain/spectral-sensitivity hybrid models (Kauramaki, Jaaskelainen, & Sams, 2007, Okamoto, 2007 #97). Instead, a more plausible theory of neural mechanisms underlying the role of top-down attention in the buildup of perceptual streams would involve top-down projections acting in conjunction with the physical stimulus as regulators or clocks for the firing patterns of neuronal populations in auditory cortex. The outcome of this bottom-up/top-down interaction might mediate changes in the response profiles of cortical neurons, via mechanisms of synaptic and receptive field plasticity which have been shown to be gated by attention; whereby attention plays a crucial role in shifting cortical circuits from one state to another depending on behavioral demands (J. B. Fritz, Elhilali, & Shamma, 2005; Polley, Steinberg, & Merzenich, 2006). We speculate that temporal patterns of neuronal firings are crucial in resolving the competition between attended and unattended objects; hence delimiting the cognitive border between different streams.

Overall, a significant outcome of this study is that it not only demonstrates a strong coupling between the measured neural representation of a signal and its perceptual manifestation, but also places the source of this coupling at the level of sensory cortex. As such, the neural representation of the percept is encoded using the feature-driven mechanisms of sensory cortex, but shaped in a sustained manner via attention-driven projections from higher-level areas. Such a framework may underlie general mechanisms of "scene" organization in any sensory modality.

Chapter 4: Stream segregation study

4.1 Introduction

In daily life situations, a vast amount of information impinges upon our ear simultaneously. Due to limited processing capacity, only a subset of the information available at the ear can be attended and processed in more detail at high level of the auditory system. Recent psychoacoustic studies on selective attention have demonstrated that human listeners can allocate attention not only to a particular location, but also to a particular feature, such as modulation rate (Grimault, Micheyl, Carlyon, & Collet, 2002), pitch (Vliegen & Oxenham, 1999) or timbre (Cusack & Roberts, 2000; for review, see B. C. J. Moore & Gockel, 2002). The auditory system extracts these features in hierarchical and parallel ways, forming various maps that are modulated by attention (Kayser, Petkov, Lippert, & Logothetis, 2005; Treisman & Gelade, 1980).

The neural correlates of the auditory spatial attention have been extensively investigated using dichotic paradigms, where subjects selectively listen to a series of tone pips in one ear and ignored concurrent tone pips in the other ear (Giraud et al., 2000; Hillyard, Hink, Schwent, & Picton, 1973; Woldorff & Hillyard, 1991). Attentional modulation effects have been ascribed to multiple levels of the auditory system including the auditory cortex (Jancke, Mirzazade, & Joni Shah, 1999), the brainstem (Lukas, 1981) and down to the cochlea (Giard, Collet, Bouchet, & Pernier, 1994). However, few studies (e.g. Bidet-Caulet et al., 2007) have examined the neural correlates of *feature*-based auditory attention, largely due to the difficulty of dissociating the neural activities specifically corresponding to distinct auditory objects. Here we tackle this issue by taking advantage of a stream segregation paradigm and the auditory Steady-State Response (aSSR), an electrophysiological signature of modulated sounds. Two concurrent streams at different amplitude modulation rates were diotically presented to human listeners. Each stream elicits an aSSR at the corresponding modulation rate. By requiring subjects selectively attend to one stream (with a specific rate) and monitoring the corresponding aSSR, the modulatory effects of rate-based selective attention can be examined.

The present study employs stimuli consisted of a two parallel tone sequences with different repetition rates, 4 Hz and 7 Hz, critical modulation rates for speech recognition (Figure 4.1). In separate tasks with identical stimulus ensembles, subjects are asked to detect a temporal jitter deviant in the stream with either the slow (4 Hz) or fast (7 Hz) repetition rate, while MEG responses are recorded. The power and phase coherence of aSSR at the modulation rates in these two tasks is compared to demonstrate that the neural representations of attended features adapt to the task. The evolution of neural representation of an attended stream is correlated with the detectabilty performance. In addition, the neural source locations of the aSSR and hemisphere asymmetry are assessed. The hypothesis is similar to the informational masking study: depending on listeners' attentional focus, the percept of auditory objects in a complex scene is differentially represented in the auditory cortex.

4.2 Methods

4.2.1 Subjects

Twenty eight subjects (13 males; mean age 26 years) participated in the MEG study. Two subjects were excluded from further analysis due to an inability to perform the tasks. All subjects were right handed (Oldfield, 1971), had normal hearing and no history of neurological disorder. The experiments were approved by the University of Maryland Institutional Review Board and written informed consent was obtained from each participant. Subjects were paid for their participation.

4.2.2 Stimulus design

The stimuli were generated using MATLAB (MathWorks Inc, Natick MA). The duration of sounds were randomly chosen from 5.25, 6.25 or 7.25 seconds uniformly to prevent the formulation of an expectation of the end of the stimulus in subjects. The sampling frequency was 8 kHz. Each stimulus contained two target notes, one repeating at 4 Hz, and the other at 7 Hz. The spectral distance between the two streams was fixed at ±8 semitones, while the specific frequencies of each stream were randomly chosen in the range 250-500 Hz in two semitone intervals. The intensity of each stream was adjusted twice to have approximately equal audibility: first roughly adjusted based on the standard equal-loudness curve (ISO-226, 2003), then fine tuned by five subjects. Tones were 75 ms in duration with a 10 ms onset and offset cosine ramps.

Twelve exemplar stimuli were generated for each of the three condition types: (i) null condition (no deviants), (ii) slow target condition (one deviant per 4 Hz stream), and (iii) fast target condition (one deviant per 7 Hz stream); a maximum of one deviant per condition was used to disambiguate subjects' true detections in the attended stream from incorrect false positives to deviants in the other stream. Each deviant was the temporal displacement of a target tone by ± 70 ms (in the 4 Hz stream) or ± 40 ms (in the 7 Hz stream) from the regular target temporal intervals. The temporal location of all deviants was approximately uniformly distributed along entire stimulus duration.

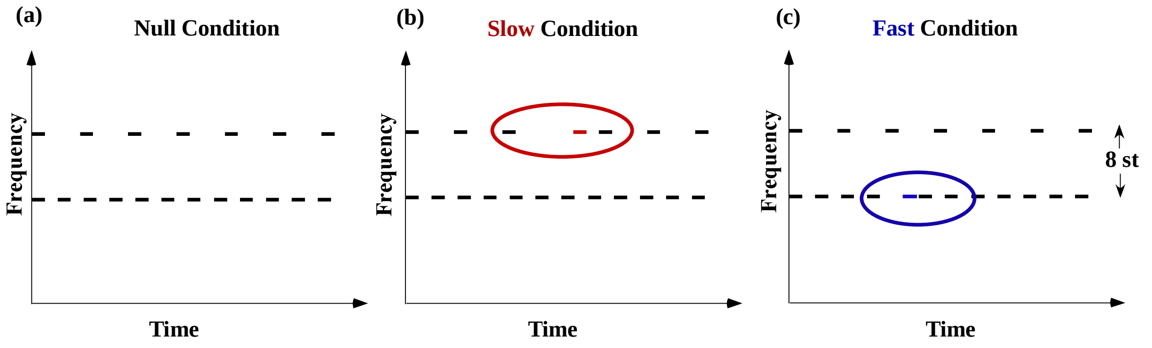

Figure 4.1: Stimulus cartoon illustration. The stimulus consists of two target notes, one repeating at 4 Hz, and the other at 7 Hz. The spectral distance between the two streams was fixed as ±8 semitones. In the T4 (T7, respectively) task subjects are instructed to track the 4 Hz (7 Hz, respectively) stream to detect a temporal jitter in that stream.

4.2.3 Experimental procedure

Subjects were placed horizontally in a dimly lit magnetically shielded room (Yokogawa Electric Corporation, Tokyo, Japan). Stimuli were presented using Presentation software (Neurobehavioral Systems, Albany, CA). The signals were delivered to the subjects' ears with 50 Ω sound tubing (E-A-RTONE 3A, Etymotic

Research, Inc), attached to E-A-RLINK foam plugs inserted into the ear-canal, and presented at a comfortable loudness of approximately 70 dB SPL.

Before the main experiment, a pre-experiment was run, where a 1 kHz, 50 ms tone pip was presented about 200 times. The inter-stimulus interval (ISI) was randomized between 750 ms and 1550 ms and subjects were instructed to count the tone pips. The aim of this task was to record the M100 response (a prominent peak approximately 100 ms after pip onset, also called N1m) used for differential source localization. The responses were checked to verify that the location and strength of neural signals fell in a normal range.

In the main experiment, subjects were presented with 36 stimuli (3 conditions × 12 exemplars) per block. The ISIs were randomly chosen to be 2800, 3000, or 3200 ms. Two blocks per task were presented, and between blocks subjects were allowed to rest but were required to stay still. The identical stimulus ensembles (including identical ISIs) were presented for two tasks. Depending on the task being performed, subjects were instructed to listen for the occurrence of temporal jitter in the slow rhythm (4 Hz, T4 task) or the fast rhythm (7 Hz, T7 task); each task deviant was present in exactly 1/3 the stimuli. Each subject performed both the T4 task and the T7 task, with task order counterbalanced across subjects. Subjects were instructed to press a button held in the right hand as soon as they heard the appropriate deviant. In the following, 'target stream' refers to the stream which is attended for that task while 'masker stream' refers to the other stream. Thus a target stream (such as the 4 Hz stream) in one task (T4 task) is a masker stream in the other task (T7 task). 'Target rate' denotes the modulation rate of the target stream.

A training block with 20 sounds was presented before each task, where each training sound was presented twice. Subjects verbally indicated the existence of the deviants and the feedback was given by the investigator. The entire session of both tasks lasted approximately 1 hour.

MEG recordings were conducted using a 160-channel whole-head system (Kanazawa Institute of Technology, Kanazawa, Japan). Its detection coils are arranged in a uniform array on a helmet-shaped surface of the bottom of the dewar, with about 25 mm between the centers of two adjacent 15.5 mm diameter coils. Sensors are configured as first order axial gradiometers with a baseline of 50 mm; their field sensitivities are 5 fT/√Hz or better in the white noise region. Three of the 160 channels are magnetometers separated from the others and used as reference channels in noise filtering methods. The magnetic signals were bandpassed between 1 Hz and 200 Hz, notch filtered at 60 Hz, and sampled at the rate of f_s = 1000 Hz. All neural channels were denoised twice with a Block Least Mean Square (LMS) adaptive filter: first using the 3 external reference channels (Ahmar & Simon, 2005), and secondly using the 2 channels with the strongest cardiac artifacts (Ahmar, Wang, & Simon, 2005).

4.2.4 Data analysis

Behavioral Performance Analysis

The ability of subjects to perform the requested task was assessed by calculating a d-prime measure of performance (Kay, 1993). For each task, we estimated the correct detection and false alarm probabilities for detecting the jitter deviants in the attending stream; converted them to normal deviates (z-scores) and computed the d-prime value.

To investigate the buildup of the target object during the corresponding task, we divided deviant trials according to 6 time windows, starting 1.25 seconds post-stimulus, to the end of the sound, with 1 second intervals. A probability of hit was then measured for each time window. Because of the temporal uncertainty in the false alarm trials, we calculated an average false alarm rate (irrespective of when the false response was issued), and combined it with the time-specific hit rate to derive a d-prime measure for each time segment.

Neural Data Analysis

After recordings were completed and noise reduction algorithms applied, the analysis epochs of each task from 1.25 seconds post-stimulus to the end of the sound were extracted and concatenated, forming a single response with duration T = 360 seconds (5 seconds × 36 sounds × 2 blocks), due to the balanced sound lengths across trials. The Discrete Fourier Transform (DFT) was applied on the single response, giving a single Fourier response of from 0 to 500 Hz with frequency resolution (Δf) 1/360 Hz.

The evoked neural responses to the two streams were characterized by the magnitude and phase of the frequency component at modulation rates (4 Hz and 7 Hz) and were used for localization and for phasor maps. The complex magnetic field strength is given by the product of the value of DFT times the sampling interval ($1/f_s$), and has units of fT/Hz. Power spectral density is calculated as the product of the inverse duration

(1/*T*) times the modulus squared of the complex magnetic field strength, and has units of fT^2/Hz. The bulk of the analysis was based on the normalized neural responses, defined to be the squared magnitude of the frequency component at target rate divided by the average squared magnitude of the frequency components ranging from 1 Hz below to 1 Hz above the target rate (excluding the component at target rate), averaged over the 20 channels with the strongest normalized neural responses. Using 10, 20 or 50 channels yielded similar findings; however, only the 20 channel analysis is reported here.

The spatial pattern of the neural responses was represented by a phasor map, a graph of the complex (magnitude and phase) magnetic field on all channels. For each channel, the length of the vector arrow is proportional to the magnitude of frequency component at target rates and the direction of the arrow represents the phase according to standard polar coordinates. Red and green contours represent the magnetic field strength projected onto the line of constant phase that maximizes the projected field's variance (Simon & Wang, 2005). The phasors are visually faded using the signal-to-noise ratio (SNR) of each channel as linear fading coefficients.

The normalized neural responses difference between T4 and T7 task was averaged across 26 subjects to characterize attention gain effect. Furthermore, to evaluate the effect of attention at across frequencies, the same analysis is done at adjacent frequency bins around the target rates (4 ± 0.25 Hz, 7 ± 0.25 Hz), which was chosen to be consistent with the frequency resolution in the following phase coherence analysis.

Phase coherence analysis is used to study the effect of attention on the synchronization between two distinct neural populations. Phase coherence between channels m and n, γ_{mn}^2 is obtained from Q trials (Srinivasan, Russell, Edelman, & Tononi, 1999) ; Here Q = 72 = 3 conditions × 12 exemplars × 2 blocks,

$$\gamma^2{}_{mn}(f) = \frac{\left| X_{mn}(f)^2 \right|}{\left\langle X_{mm}(f) \right\rangle \left\langle X_{nn}(f) \right\rangle}$$

where, $X_{mn}(f)$ is average cross spectrum between channel m and channel n, $X_{mn}(f)$ is average power spectrum of the individual channel,

$$X_{mn}(f) = \frac{1}{Q} \sum_{q=1}^{Q} F_{mq}(f) F_{nq}(f)^*$$

where $F_{mq}(f)$ is the Fourier transform of the qth trial of channel m at frequency f. Each trial is extracted from 1.25 to 5.25 seconds post-stimulus, giving a frequency resolution 0.25 Hz. A coherence value of 1 indicates that the two channels maintain the same phase difference on every trial, whereas a coherence value near 0 indicates a random phase difference across trials. The coherence difference between T4 task and masker task was computed for every channel pair. The SEM ε_{mn} was constructed to identify robust coherence change (Srinivasan, Russell, Edelman, & Tononi, 1999).

$$\varepsilon_{mn} = \sqrt{\frac{2}{Q}} \left(\frac{1 - \gamma^2{}_{mn}}{\left| \gamma_{mn} \right|} \right)$$

To analyze the modulation of phase coherence in the auditory cortex, only each subject's 20 channels with the strongest normalized neural response at target rates are included in further analysis. In addition, to exclude artificial coherence resulting from volume conduction effects on the extracranial magnetic field and to measure only phase correlation between distinct populations of neurons, only long distance channel pairs (channel separation > 100 mm) were included (Srinivasan, Russell, Edelman, & Tononi, 1999). The difference between number of channel pairs with robust increased coherence and channel pairs with decreased coherence is normalized over the total number of long range channel pairs. Finally, to evaluate synchronizations at other frequencies, the coherence change measurement is obtained at adjacent frequency bins (4 ± 0.25 Hz, 7 ± 0.25 Hz).

To investigate the possibility of hemispheric bias, the 20 channels with the strongest normalized neural response at target rates are chosen from the left and right hemispheres respectively to represent the overall neural activity of each hemisphere. Neural responses averaged across the 20 channels are subtracted across hemispheres for each task and for each subject. Using 10, 20 up to 60 channels yielded similar findings; however, only the 20 channel analysis is reported here.

To investigate the buildup of the representation of target stream, the responses at target rate were divided temporally: the analysis epochs were divided into four temporal segments with 1000 ms duration each and corresponding segments were concatenated across epochs. The segment duration 1000 ms was used to be commensurate with both 4 Hz and 7 Hz. The first segment began at 1250 ms post stimulus.

Behavioral vs. Neural Correlation and Bootstrap Analysis

We correlated the psychometric and neurometric curves for the deviant detection buildup. To match the range of values from the neural and behavioral data, we scaled the neural data (i.e. the normalized responses to target rate) by a factor of 2.5. The behavioral curves for each subject were then grouped to match the sampling rate of the neural data. Subsequently, these two curves were fitted by a first-order polynomial to derive the slope relating the two functions. The across subject slopes were then combined using circular statistics to yield an angular mean for each task (Fisher, 1993).

We performed a bootstrap procedure in order to confirm the positive correlation between the neurometric and psychometric functions in the target task. We followed a balanced bootstrap sampling procedure (Efron & Tibshirani, 1993) by randomly selecting 26 subjects with replacement and computing their angular sample mean and repeating this process 1000 times. The procedure was controlled to ensure that all subjects appear the same number of times over all 1000 bootstrap samplings. Confidence measures were then derived from the bootstrap statistics (Efron & Tibshirani, 1993).

Neural Source Localization

Source localization for the M100 response was obtained by calculating the current-equivalent dipole best fitting the magnetic field configuration at the M100 peak, in each hemisphere. Source localization for the neural response to target stream was obtained by calculating the complex current-equivalent dipole best fitting the complex magnetic field configuration at target rates, in each hemisphere (Simon & Wang, 2005). Only channels with SNR > 4 were used in the fitting. Significance of the relative displacement between the M100 and aSSR dipole sources were determined by a two-tailed paired t-test in each of three dimensions: lateral/medial, anterior/posterior, and superior/inferior.

4.3 Results

4.3.1 Enhanced power at target rates

Depending on listeners' attentional focus, the neural representations of the streams mirrors the percept of a scene with two distinct streams. During the performance of the T4 task (mean d-prime: 2.9) the rhythm of the slow stream emerges as a strong 4 Hz component in the neural signal of an individual subject (Figure 4.2a, left panel). In contrast, during the performance of the T7 task (mean d-prime: 1.8) the cortical response entrained at 4 Hz is relatively suppressed (Figure 4.2a, right panel). This modulatory effect is reversed for the cortical representations of the fast stream: the neural response at 7 Hz is stronger in the T7 task than the T4 task (Figure 4.2b). This differential activation confirms our previous findings in the informational masking study that task-dependent attention *sustained* modulates the neural representation of a single acoustic stimulus, much like visual attention.

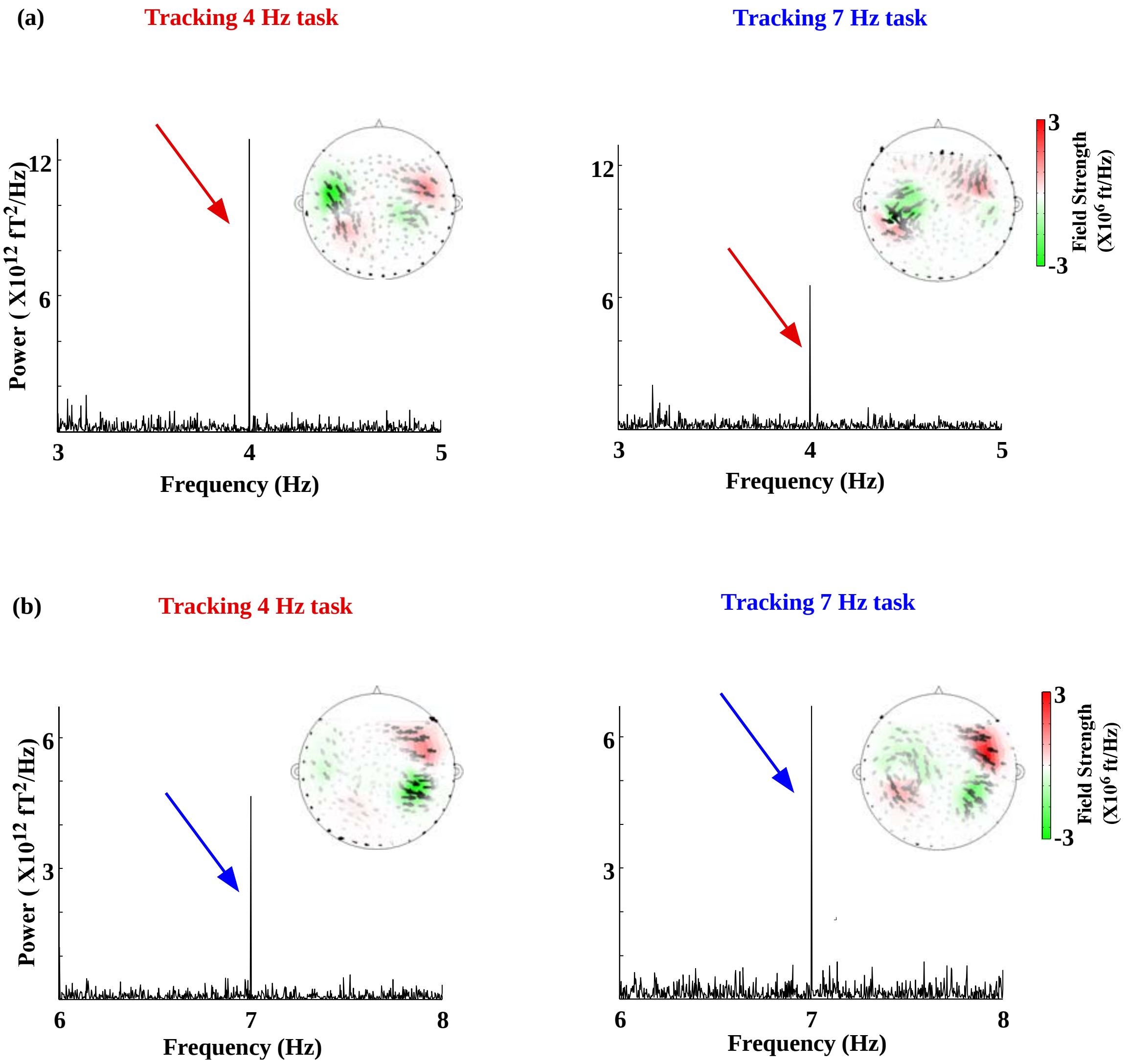

Figure 4.2: a) Power spectral density of the 4 Hz aSSR for a single subject in T4 (left) and T7 (right) tasks, averaged over twenty channels. Insets: The MEG magnetic field distributions of the 4 Hz rhythm response component. Red and green contours represent the target magnetic field strength projected onto a line with constant phase. b) Power spectral density of the 7 Hz aSSR for the same subject in T4 (left) and T7 (right) tasks.

The MEG magnetic field distributions of the target rate response component, examples of which are shown at the inset of graphs in Figure 4.2a and Figure 4.2b, reveal the stereotypical pattern for neural activity originating separately in left and right auditory cortex. The neural sources of all the target rhythm response components with sufficiently high signal-to-noise ratio originate in auditory cortex (Simon & Wang, 2005). The neural source's mean displacement from the source of the auditory M100 response (Naatanen & Picton, 1987) was calculated for each hemisphere and each target rate. No significant difference of displacement was found between hemisphere and target rate. Thus these displacement are combined, yielding a significant displacement ($p = 0.016$) of 19 ± 6 mm in the anterior direction. Assuming an M100 origin of *planum temporale*, this is consistent with an origin for the neural response to the target rhythm in Heschl's gyrus, the site of core auditory cortex.

This task-depended attentional effect is confirmed in the population of 26 subjects (Figure 4.3a). The neural response change at target rates across tasks were averaged across all subjects. A significant positive (bootstrap across subjects, $p < 10^{-3}$) change at the 4Hz aSSR and a significant negative (bootstrap across subjects, $p < 0.002$) change at the 7 Hz aSSR are observed, reflecting an enhanced phase-locked, sustained activity when subjects' attention is directed towards the target stream. In contrast, there is no significant change in normalized neural response at adjacent frequencies (4 ± 0.25 Hz, 7 ± 0.25 Hz), confirming that this feature-based selective attention precisely modulates the cortical representation of the specific feature, but not overall neural activities.

4.3.2 Enhanced long-distance coherence at target rates

Feature-based attention enhances long range synchronization at the target rates. Phase coherence between distant MEG channels is used to characterize synchronization of underlying neural activities. The difference between number of long range channel pairs with robust increased coherence in the attended task and channel pairs with decreased coherence is normalized over the total number of long range channel pairs (Figure 4.3b). Approximately 15% more channel pairs show an enhanced synchronization at target rates (bootstrap across subjects, $p < 0.008$ at 4 Hz and $p < 0.003$ at 7 Hz). In contrast, there is no significant change in phase coherence at adjacent frequencies (4 ± 0.25 Hz, 7 ± 0.25 Hz). The channel pairs with robust coherence difference at target rates in one subject is shown by red lines (denoting an increased coherence) and blue lines (denoting a decreased coherence), plotted on the contour map of normalized neural response at target rates (Figure 4.3c). The coherence change is distributed both within and across hemispheres.

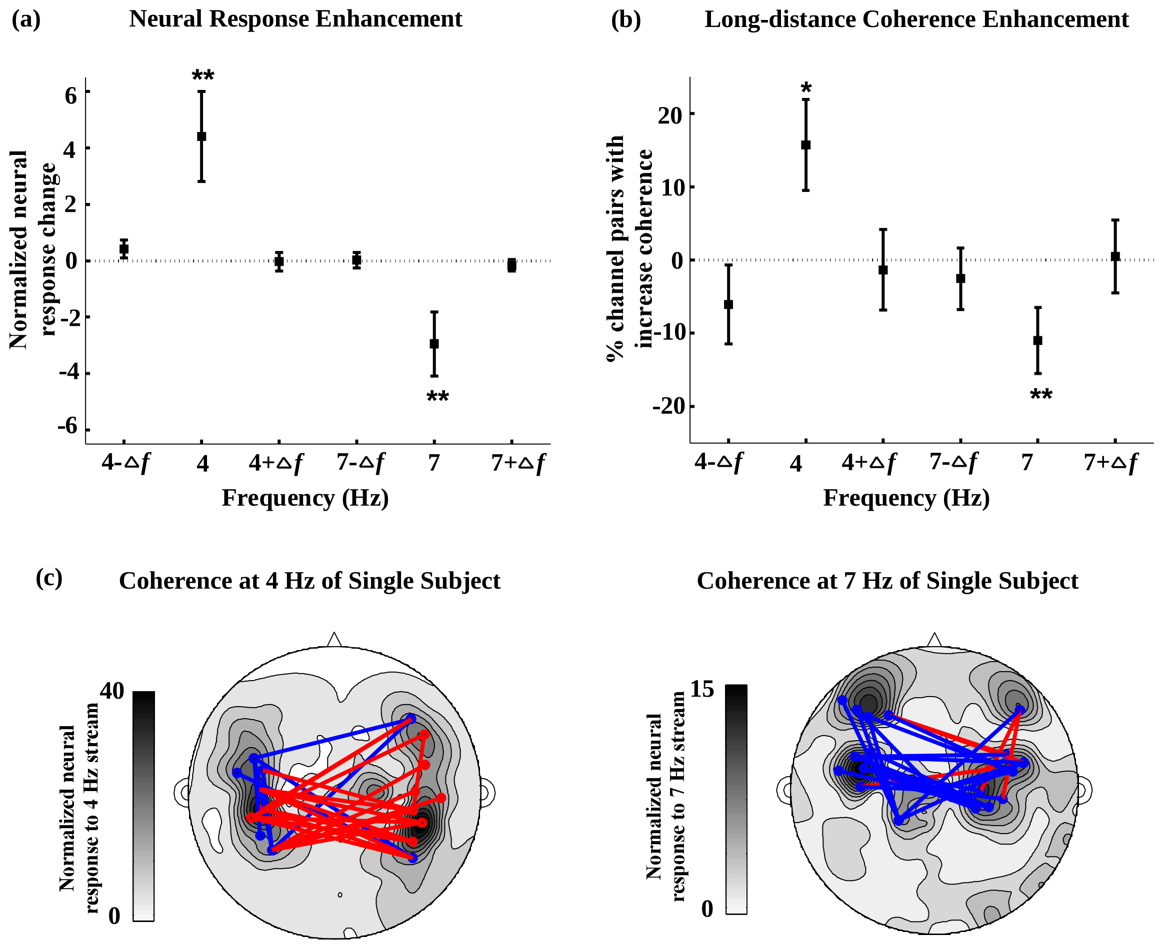

Figure 4.3: Power and phase enhancement during the attended task. a) Normalized neural responses difference between T4 task and T7 task shows enhancement exclusively at target rates (4 Hz in T4 task, 7 Hz in T7 task). Error bars represent standard error. The asterisk at 4Hz and 7 Hz shows that only that particular frequencies yields a statistically significant enhancement. Normalized neural response is computed as the ratio of the neural response power at target rates (4 Hz, 7 Hz) to the average power of the background neural activity (from 3–5 Hz, from 6-8 Hz; see METHODS), averaged over the 20 channels with the strongest ratio. The normalized neural response change is calculated at target rates (4 Hz, 7 Hz) and adjacent frequency bins. b) Phase coherence difference between T4 task and T7 task shows enhancement exclusively at target rates (4 Hz in T4 task, 7 Hz in T7 task). Error bars represent standard error. The asterisk at 4Hz and 7 Hz shows that only that particular frequencies yield a statistically significant enhancement. Phase coherence is computed as the ratio of average cross spectrum between each channel pair, and average power spectrum of the individual channels (see METHODS). The difference

between number of channel pairs with robust increased coherence and channel pairs with decreased coherence is normalized over the total number of long range channel pairs. The phase coherence change is calculated at target rates (4 Hz, 7 Hz) and adjacent frequency bins. c) Channel pairs with robust coherence difference at target rates for single subject, overlaid on the contour map of normalized neural response at target rates. The channel pairs with increased (respectively, decreased) coherence at target rates is shown by red (respectively, blue) lines. The channel pairs with robust coherence change are distributed both within and across hemispheres.

4.3.3 Build-up of power at target rates

The perceptual detectability build-up of the 4 Hz target stream is highly correlated with the neural buildup of the signal representation, but decoupled from the actual constant acoustics (Figure 4.4a), consistent with previous findings of buildup in the informational masking study. Underlying mechanisms have been suggested analogous to those employed in auditory streaming and object formation. The MEG magnetic field distributions of the 4 Hz target response in Figure 4.4a (insets), showing the stereotypical pattern of neural activity originating separately in left and right auditory cortex, illustrate the changing strength and changing pattern of the neural activity over time in an individual subject. The correlation between the psychometric and neurometric curves over time is confirmed by a bootstrap analysis on a per-subject basis. The slope correlating the d-prime and neural response curves for each subject yield a mean positive slope angle of 22º; statistical significance is confirmed by a bootstrap mean of 21º, with the 5th to 95th confidence intervals falling within the upper right quadrant (Figure 4.4b).

The behavioral detectability of the 7 Hz stream shows an initial build-up with decay afterwards (Figure 4.4c). The neural representation of the 7 Hz is approximately constant over time. The difference of behavioral and neural responses between the 4 Hz and 7 Hz stream is discussed in below section 4.4.

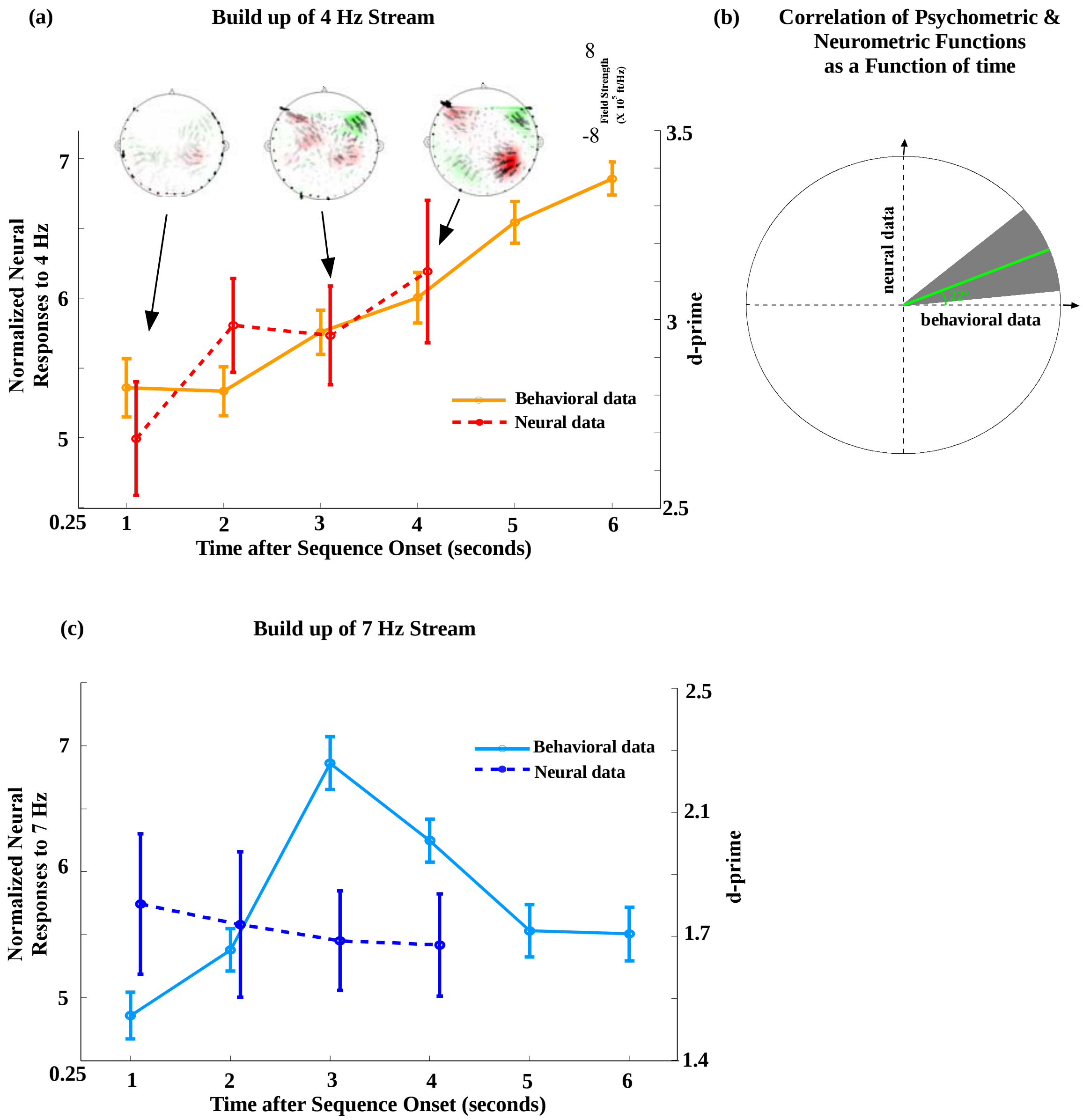

Figure 4.4: Buildup over time of behavioral and neural responses of target streams. a) Normalized neural response to the 4 Hz stream, and behavioral performance, as a function of time in T4 task averaged over subjects. Error bars represent standard error. Insets: The MEG magnetic field distributions of the 4 Hz target response for a single subject at representative moments in time. b) Correlation of behavioral in T4

Task and neural responses to the 4 Hz stream as a function of time. The ratio of the neural to behavioral response trends as a function of time, interpreted as a slope angle, is averaged across subjects yielding a mean slope angle of 22° (yellow line). Bootstrap estimates (overlaying green line) and the 95% confidence intervals (gray background) confirm the positive correlation between the psychometric and neurometric buildup curves. c) Normalized neural response to the 7 Hz stream, and behavioral performance, as a function of time in T7 task averaged over subjects. Error bars represent standard error.

4.3.4 Hemispheric asymmetry

We observe a task-dependent hemispheric asymmetry in the representation of the neural response at target rates. During the attended task, the right hemisphere shows a greater normalized neural response at target rates than the left hemisphere (bootstrap across subjects, $p < 0.038$ at 4 Hz and $p < 0.001$ at 7 Hz); during the ignored task the right hemisphere dominance is observed at 7 Hz (bootstrap across subjects, $p < 0.025$), but not at 4 Hz (bootstrap across subjects, $p < 0.07$) (Figure 4.5).

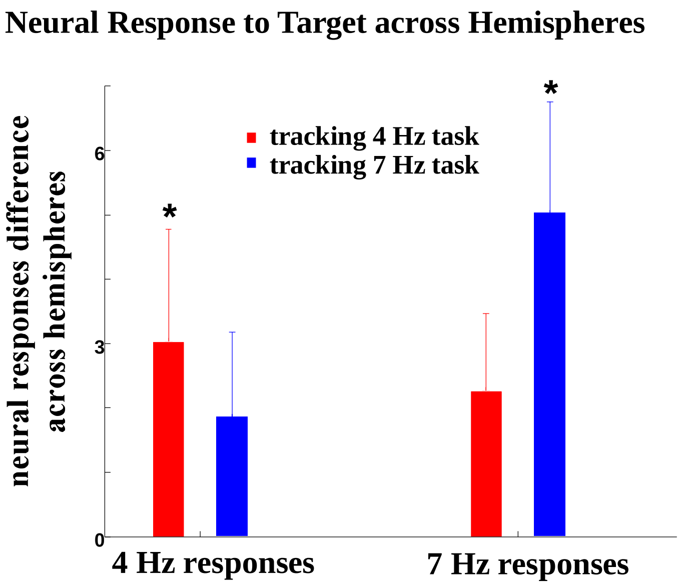

Figure 4.5: Neural responses to target rates across hemispheres. The 20 channels with the strongest normalized neural response at target rates are chosen from left and right hemisphere respectively to represent the overall neural activity of each hemisphere. Neural responses averaged across the 20 channels are subtracted across hemispheres for each task and for all subjects. Error bars represent standard error. The figure shows a task-dependent hemispheric asymmetry in the representation of the neural response at target rates

4.4 Discussion

The present study investigates the modulatory effects of feature-based attention using a stream segregation paradigm. The results presented here confirm and extend our findings in the informational masking study that feature-based attention modulates cortical representations of the feature, by boosting both power and coherence at modulation rate of the attended stream, relative to levels seen when attention is focused away from this rate. The change in cortical representations can be reversed in a short time-scale (several minutes) by various behavioral goals. Moreover, the neural representation of an attended object with a slow modulation builds up over time (several seconds) in a similar manner as the buildup of listeners' detectability performance. The neural source locations of aSSR are anterior relative to the source of M100, consistent with an anterior 'what' pathway. The neural responses for the target stream are more activated in right hemisphere during the attended task.

An increase in power of neural responses at the MEG auditory sensors during the attended task presumably reflects local synchronization of underlying neural population (Hamalainen, Hari, Ilmoniemi, Knuutila, & Lounasmaa, 1993). Slow temporal modulation (< 20 Hz) is represented by temporal coding in neurons with band-pass modulation transfer functions (MTF). Thus, increased power of neural response at target rate could be either a consequence of a shift of the best modulation frequency (BMF) to the target rate in neurons, or a consequence of improved precision of the temporal coding within the neuron population with BMF equal to target rate. Both situations reflect a short-term plasticity of the cortical neurons. Parallel with task-induced plasticity in a spectral response feature seen in single-neuron recording (J. Fritz, Shamma, Elhilali, & Klein, 2003), here we show task-induced plasticity of the temporal modulation feature. Such feature specific modulation of neuronal receptive fields by attention may facilitate adjustment and calibration of the perceptual system based on the particular acoustic environment and task requirements.

An increase in coherence between distant MEG sensors reflects increased synchronization between distinct neural populations located in different brain region. An increase in long-distance coherence has been associated with the coherent (Gaetz, Weinberg, Rzempoluck, & Jantzen, 1998) and conscious percepts (Srinivasan, Russell, Edelman, & Tononi, 1999) in visual modality. The present results provide the first demonstration that selective attention increases synchronization between distinct neuron populations. The comparable results across modalities suggest that the global synchronization may be a general neural expression of cognitive functions in multiple sensory domains. Synchronization may serve as a tag of relatedness and promotes the saliency of the perceived feature (here, modulation rate).

We speculate the modulatory effects (increased local and global synchronizations) of attention are asymmetric for stimuli with different saliency. Because humans are more sensitive to slow modulations, as reflecting by the low-pass shape of behavioral MTF, the 4 Hz stream is a more salient and stronger distracter in the T7 task. Consequently, we observe a behavioral asymmetry (d-prime, 2.9 vs. 1.7), correlating with an asymmetry of modulatory effect (mean of change in power, 4.40 vs. 2.94; mean of change in coherence, 15% vs. 11 %). The asymmetry effect may be generalized to any scene consisted of competitive objects. In a rivalry situation, difference in feature values and the saliency of each stream may affect the competitive selection and automatic bottom-up filtering of objects, thus change the ease of resolving competitive objects and the accuracy of identifying the target object (Knudsen, 2007).

The target's perceptual detectability of the 4 Hz stream builds up correlates strongly with the target representation's neural build-up, consistent with the findings in the informational masking study. This progressively enhanced representation of the attended stream at the system level is in contrast with the gradually decreased response observed in single neuron, implying that different coding mechanisms are employed at different levels. The build-up effect may be explained by the operations of top-down attention. At the onset of stimuli, multiple objects in a complex scene are processed and compared with the template stored in the working memory which may locate in prefrontal cortex (PFC). A bias signal in favor of the matched stream is sent back to sensory cortex to boost the representations of the target. The bi-directional information flow between the PFC and sensory cortex may go back and forth several times to evaluate and confirm a correct operation, resulting in a gradually enhanced neural representation of the target stream and a progressively improved detectability of the target stream.

The behavioral evolution of the 7 Hz stream includes an initial build-up and decay afterwards. Two factors may be responsible for the discrepancy between the evolution of the 4 Hz and the 7 Hz stream. One factor could be the different durations of jitters in the two streams. A short jitter (40 ms vs. 70 ms), which is harder to detect, was used in the T7 task. However, we argue that the factor of jitter is not, at least not the only, reason for this discrepancy. Because the greater difficulty in the T7 task would uniformly shift down the target detectability curve, relative to the one of the 4 Hz stream; i.e., it would not change the shape of the curve. We speculate the fast repetition rate contribute to this discrepancy. Because streaming occurs faster for tone sequences with fast repetition rate, building up of the 7 Hz stream may take about half the time as it takes to build-up of the 4 Hz stream. On the other hand, the comparison and matching processes (necessary components in top-down attention) in working memory update more frequently due to the fast repetition rate of stimuli. Therefore, approaching the end of the stimuli, a fatigue factor kicks in and drags the behavioral curve down gradually. Note that in our study, participants were asked to maintain attention on a particular stream rather than to simply count the number of streams, as in traditional streaming experiments.

In a recent neuroimaging study (Ahveninen et al., 2006) locations of feature-based and space-based attention were compared. A posterior M100 component was proposed to account for space-based attention, while an anterior M100 component was proposed to account for feature-based attention, consistent with the putative “what” and “where” pathway. Our study find the source location of MEG responses to modulate rate in more anterior region of the auditory cortex relative to the source location of M100, confirming that feature-analysis processes are located in the anterior auditory cortex.

Right hemisphere dominance is observed in this study, where two streams are presented concurrently, forming a sound rivalry situation. To fulfill the detection task, it is necessary to focus on the target stream while suppressing the masker stream. The right hemisphere have been suggested as processing relevant information and suppressing irrelevant information, respectively (Alcaini, Giard, Echallier, & Pernier, 1995), hence a strong activation of right hemisphere is expected in sound rivalry situation.

An increase in neural representations of target streams may be explained by a release from suppression or an enhancement of target streams or both. Hence the current paradigm does not differentiate gain model from sharpening model (J. B. Fritz, Elhilali, & Shamma, 2005; Okamoto, Stracke, Wolters, Schmael, & Pantev, 2007). Future studies may include a passively listening condition and an analysis on the neural responses at interaction components (3 Hz or 11 Hz here, 7 Hz +/- 4 Hz) to determine the interaction and the temporal order between enhancement and suppression.

In summary, these results demonstrate that the neural representation of the percept is encoded using the feature-driven mechanisms of sensory cortex, but shaped in a sustained manner via attention-driven projections from higher-level areas. This adaptive neural representations occur on multiple time scales (seconds vs. minutes) and multiple spatial scales (local vs. global synchronization). Such multiple resolutions of adaptation may underlie general mechanisms of "scene" organization in any sensory modality and may contribute to our highly adaptive behaviors.

Chapter 5: Summary and future work

5.1 Summary

Hearing has critical social and survival value for organisms. Although seemingly effortless, the auditory system performs incredibly complex analyses on incoming sounds. Emerging evidence suggests multiple features are extracted and encoded in the auditory system. The goal of the current work has been to investigate the cortical representations of temporal modulation under different stimulus and behavioral contexts.

In chapter 2, we focused on representations of compound modulations. It is found that rate separation between these two modulations has a significant effect on the pattern of neural activity in neuron ensembles. Interaction components are evoked when rate separation is less than 3 Hz. In contrast, no interactive components are observed for large rate separation (> 17 Hz). Our results provide physiological evidence to support the existence of modulation filter banks whose filter bandwidths are band-limited (with respect to modulation rates). In addition, compared with high modulation rates (> 20 Hz), slow modulation rates generate larger aSSR amplitudes, forming an overall low-pass shape. Furthermore, and crucially, aSSR at the lowest frequency, regardless of whether it is elicited by the stimulus directly or indirectly, correspond with the perceived loudness fluctuation of the dual-modulated stimulus. Based on these observations, a simple model of modulation filter banks is proposed (Figure 5.1). Note that the magnitude of each modulation filter decreases with increasing rate, consistent with psychoacoustic observations in humans. The filter centered around 20 Hz has a bandwidth of at least 3 Hz. In converse, the streaming segregation study shows 3 Hz separation is enough for segregation of the 4 Hz stream and the 7 Hz stream, suggesting a bandwidth of less than 3 Hz for filters with a center rate around 4 Hz. Therefore, we speculate that the bandwidth of each filter is proportional to the center rate of the filter, i.e., a filter with a low center rate has a narrower bandwidth. The plasticity of modulation filter banks is depicted by a red curve and arrow and is discussed in the context of study two and study three.

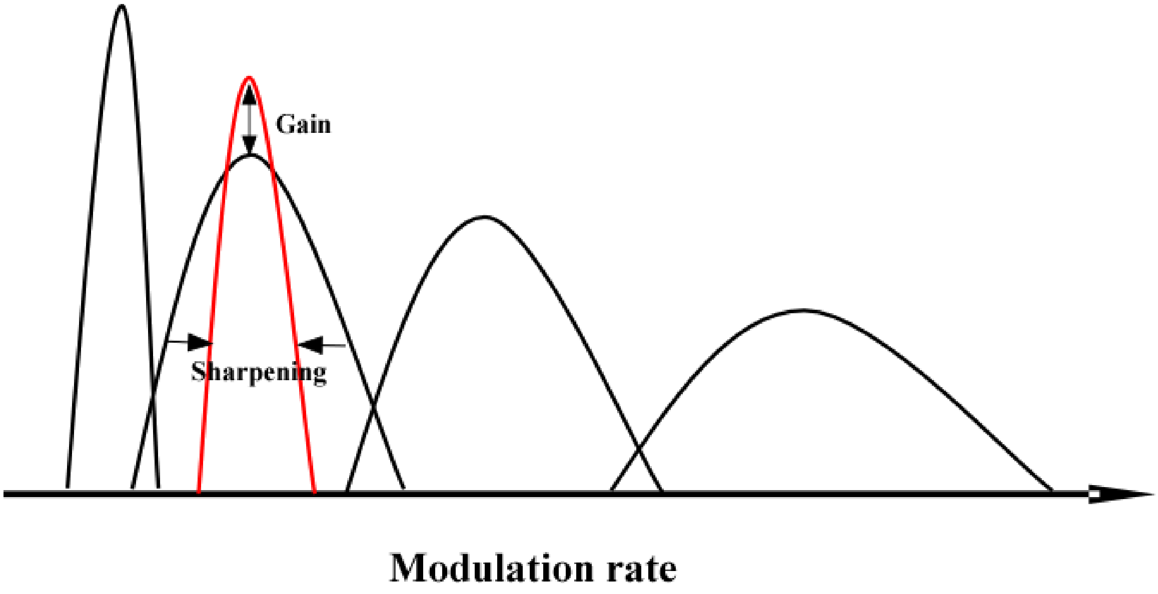

Figure 5.1: Schematic representation of modulation filter banks. The bandwidth of each filter is increased with increasing center frequency. The overall shape is low-pass. Gain and sharpening effects of attention are represents by a red curve.

Chapter 3 explored the effects of task-driven and stimulus-driven attention on cortical representations of temporal modulation. The informational masking paradigm is used to represent a complex auditory scene. Subjects' attention is manipulated to different features of an auditory scene in a simultaneous behavioral-neurophysiological manner. Our experimental results reveal that target-focused attention correlates with a sustained increase in the neural target representation, beyond auditory attention's well-known transient effects. The enhancement originates in auditory cortex and co-vary with both the behavioral state appropriate to the task and the bottom-up saliency of the target. Furthermore, the target's perceptual detectability improves over time, correlating strongly with the target representation's neural buildup. Finally, a left-hemisphere dominance is observed in the attended task.

Neural activities at the system level should be related with the activity at the single-neuron level. An increase in power of the aSSR at the majority of MEG auditory sensors during an attended task presumably reflects the local synchronization of an underlying neural population. Hence, a gain effect is observed at the level both of neuron ensembles and the single neuron (red arrow in Figure 5.1). An enhancement in phase coherence between long-distance channel pairs reflects increased long-distance coherence at the system level. The information about long-distance coherence remains unseen at the single-neuron level, due to the spatial limitations of this technique.

In chapter 4, we employed a stream segregation paradigm, in which two streams were presented concurrently, to investigate the modulatory effects of feature-based attention. Results of this study confirm and extend the findings in chapter 3 that feature-based attention modulates sensory representations of the feature, by boosting both power and coherence at the modulation rate of the attended stream. This study also shows that the neural representation of an attended signal builds up over time in a manner analogous to the buildup of listeners' detectability performance. Finally, hemispheric lateralization is observed: the right hemisphere is more activated in the attended task. The significance of these findings is discussed in the following sections. The commentary is given to discuss the similarity and discrepancies across these two studies.

5.1.1 Local synchronization

In both studies, the power at the modulation rate of the attended stream is enhanced at bilateral channels, reflecting that the local synchronization of neurons at the sensory cortex is improved in a sustained manner, beyond the well-known transient effect observed in most other physiology studies. The question remains as to what neural mechanisms cause the sustained effect. Does the effect result from the attention process at the high level continually interacting with sensory representations at the low level? Or from a continuous bias signal which is itself located in the sensory cortex but is triggered by a transient control signal from the attention process? Is the attention process located in a distinct anatomical area? If it is, how do the attention neurons encode command messages, and how does the sensory cortex read out these messages? These questions directly address the essence of attention, instead of various phenomena that are associated with attention. A number of lesion studies suggest a central role of the prefrontal cortex (PFC) in attention. Lesions in the PFC cause general deficits of attention, including neglect, schizophrenia, and closed head injury (Duncan, Emslie, Williams, Johnson, & Freer, 1996). To further understand the neural correlates of attention and develop new approaches to these pathologies, future studies should examine activity in the PFC and relate it to activity in sensory cortex.

5.1.2 Global synchronization

Long-distance coherence at the modulation rate of the attended stream is enhanced intra-hemispherically and inter-hemispherically, demonstrating an increased global synchronization between different brain regions. An increase in coherence has been associated with coherent and/or conscious perception in the visual modality (Srinivasan, Russell, Edelman, & Tononi, 1999). Our studies first demonstrated the tight coupling between global coherence and attention in the auditory domain. The fact that our results are compatible with data about conscious perception in visual studies suggests at least two points. First, global synchronization may be a general neural coding strategy of cognitive functions in multiple domains. Synchronization serves as a tag of relatedness and promotes the saliency of the feature (in our case, the attended modulation rate) reaching the consciousness, which in turn directs bias signals in favor of the attended feature, forming a recurrent loop. It is noteworthy that synchrony here is different from synchrony in binding theory, where it refers to the process of linking different features (such as color, shape) to form a coherent object. The observed global synchronization can be explained either by reciprocal connections between distributed neurons linked via cortico-cortical fibers or by reciprocal connections between cortical and subcortical areas. The latter connection would increase synchronized input to cortical neurons, and thus necessarily enhance synchronization between cortical neurons. Further experiments are necessary to dissociate these two potential mechanisms.

Coherence measurement provides a way to address the relationship between sensory cortex and the PFC. Coherence between aSSR at sensory channels and functional oscillations (such as the theta and gamma oscillations) at frontal channels might increase during an attended status but break down in an ignored condition. In addition, the extent of synchronization may relate to the amount of attention required in various tasks. The direction of information flow can be determined to evaluate the causality between the two areas (Brancucci et al., 2005). The front channels, which have maximum coherence with sensory channels, might define the anatomical location of attention, the 'command center'.

Second, consciousness and attention are related but distinct processes. Often though, the two concepts are thought to be equivalent, in that we are conscious of what we attend to. Likewise, when we withdraw attention from an object, we lose consciousness of its attributes. However, recently it has been argued that attention and consciousness are distinct processes (Crick & Koch 2003). Here we take a view that consciousness is equivalent with bottom-up attention (pop-out) but largely different from top-down attention. We argue two important differences between consciousness and top-down attention, maintenance of a working memory and a top-down control signal. At any moment, a salient stream accesses our consciousness (it pops out). In the next moment, another stream with the same saliency may replace the current one to access our consciousness. No working memory is maintained and no control signal (in favor of any particular stream) is being sent back to the sensory cortex. In other words, consciousness is a one-way process. However, for top-down attention, a working memory must be maintained to store a template. Moreover, a bias signal in favor of the matched stream (with respect to the template) must be sent back to the sensory cortex, creating a loop. Perhaps it is the non-maintenance of working memory that makes us feel consciousness is so effortless (it pops out!), while paying attention to something is tiresome due to the necessity of maintaining working memory. Having identified the similarities (power and coherence enhancement) between top-down attention and consciousness, the next step is to search for their different neural mechanisms.

An interesting question emerges from the discussion of the distinction between consciousness and attention. What are their respective functional roles? Why does our brain need both of them? Before answering these questions, let us revisit the challenging task faced by the brain, which was described at the beginning of the thesis. The brain must prioritize the relevant tasks at hand while maintaining the capability to react quickly to unexpected events. We proposed that top-down attention functions to process task-related information while unexpected events are caught by bottom-up attention. Whether the bottom-up attention will override the top-down attention to take over the limited computational resources is determined by the emergence of the unexpected events. An analogy is the dual tasks performed by the operating system of a computer. Most of the time, the foreground task is running and using all kinds of resources, such as CPU, memory. However, a second event might be picked up by background. If the priority level of the new event is higher than the current one, the current one will be suspended and the new one starts. In explaining the functions of consciousness, Crick and Koch (1995) remarked:

"Our second assumption is based on the broad idea of the biological usefulness of [visual] awareness (or, strictly, of its neural correlate). This is to produce the best current interpretation of the [visual] scene, in the light of past experience either of

ourselves or of our ancestors (embodied in our genes), and to make it available, for a sufficient time, to the parts of the brain that contemplate, plan and execute voluntary motor outputs (of one sort or another).”

5.1.3 Build-up of the percepts

In both studies, the buildup of the perceptual detectability of the attended 4 Hz stream correlates strongly with the build-up of neural representations of the stream. However, the behavioral/neural evolution of the 7 Hz stream does not show build-up. In the following, we discuss the neural-behavioral evolution of 4 Hz and 7 Hz streams.

The neural mechanism behind psychoacoustic build-up has been suggested to be ‘adaptation’ or ‘habituation’ of single neural responses over time (Fishman, Reser, Arezzo, & Steinschneider, 2001). Our studies offer the first demonstration of the build-up of the neural representation at the system level. Surprisingly, the neural evolution is not adapted per se, but gets stronger over time. The progressively enhanced representation of the attended stream might be accounted for by the function of top-down attention. At the onset of sound stimuli, multiple streams in the complex scene are processed and compared with the template stored in the working memory. After the comparison, a bias signal in favor of the matched stream is sent back to sensory cortex to boost the representations of the target stream. The bi-directional information flow between the PFC and sensory cortex may go back and forth several times to evaluate and confirm a correct operation, resulting in a gradually enhanced neural representation of the target stream and progressively improved detectability of the target stream. Furthermore, the buildup of neural responses over time is seen only when integrated over several periods of the target rhythm, but not for individual periods. The phenomenon suggests that the enhanced representations are more likely from the result of improved precision of temporal coding instead of a general increase of neural spiking.

The behavioral evolution of 7 Hz stream can be described as an initial build-up with decay afterwards. Two factors may be responsible for the discrepancy between the evolution of the 4 Hz and the 7 Hz stream. One factor could be the different durations of jitters in the two streams. Short jitter (40ms vs. 70ms), which is harder to detect, was used in tracking 7Hz task. However, we argue that the factor of jitter is not, at least not the only reason for this discrepancy. Because the greater difficulty in tracking the 7 Hz task would uniformly shift down the target detectability curve, relative to the detectability curve of the 4 Hz stream; i.e., it would not change the shape of the curve. We speculate that the fast repetition rate contributes to this discrepancy. Because streaming occurs faster for tone sequences with a fast repetition rate, building up the 7 Hz stream takes about half the time as it takes to build-up the 4 Hz stream. On the other hand, the comparison and matching processes (necessary components in top-down attention) in working memory need to be updated more frequently due to the fast repetition rate. Therefore, approaching the end of the tone sequence, a fatigue factor kicks in and drags the behavioral curve down gradually. The fewer samples and the large error bar of the neural response hinder detailed interpretations. Note that the first point in the curve is 1.25 seconds after the stimulus onset, by which point the buildup might have occurred and the fatigue kicked in.

5.1.4 The brain locus of attention

An anatomical consistency across the two studies is that relative to M100, the locations of the attended aSSR are more anterior, consistent with the putative anterior 'what' pathway (Ungerleider & Mishkin, 1982). Considering that both studies require feature-based attention instead of space-based attention, an anterior locus analyzing features and identifying objects is expected. In converse, in a study of space-based attention, a more posterior locus is expected, as suggested by a 'where' pathway.

Different forms of hemispheric specialization are observed in these two studies. Left-hemisphere dominance during an attended task is shown in the informational masking study, while right-hemisphere dominance is observed in the stream segregation study. The differences may be explained by subjects' use of different strategies. In the case of informational masking a search strategy is involved to parse the target stream from random notes, while in the case of stream segregation two streams are presented concurrently, forming a sound rivalry situation. To fulfill the task in the second study, therefore, it is necessary to focus on the target stream while suppressing the irrelevant stream. The functions of the left hemisphere and the right hemisphere have been suggested to be processing relevant information and suppressing irrelevant information, respectively (Alcaini, Giard, Echallier, & Pernier, 1995).

A further examination of the nature of these two studies yields some interesting points. In informational masking studies, a salient target is extracted from the background that does not have any regularity in either spectral temporal dimension. The irregularity may be interpreted by the brain as signaling the irrelevance of the corresponding event.

Therefore the event will not be caught by the wandering bottom-up attention. The weak saliency of the background facilitates the comparison and matching process of the target stream, generating a good performance of detecting deviants in the target. But in the second study, both sequences are regular. Even when we intend to focus on one stream, the other one is always picked by the bottom-up attention due to its regularity. To fulfill the task at hand, top-down process must always override bottom-up attention. The constant overriding is opposes to the situation in real life, where bottom-up attention can override and interrupt top-down attention, so that an emergency can be handled in time. This opposition may explain the overall inferior task performance in the streaming study

5.2 Future prospects

5.2.1 Methodological considerations

How can the MEG technique be maximally utilized? Should we interpret MEG responses mainly in the time or the spectral or the source domain? These questions are still hotly debated. The auditory steady state response provides a valuable tool to dissociate the activity of neural populations to distinct modulations in sound stimuli. The power, phase, spatial distribution and neural source location of aSSR are related to stimulus and behavioral contexts. This frame of experimental design and data analysis may be extended to several other paradigms, such as tone sequences with triplet ABA- and sine wave speech, where each stream might have different frequency modulation rates. The approach presented here is potentially of great importance for our understanding of ASA.

Despite these advantages, the aSSR method is intrinsically limited, in that it may only measure modulation rates below 100Hz. Modulation rates above 100 Hz are presumably entirely encoded by rate coding, which is invisible from the MEG point of view. At higher rates of modulation, although subjects still perceive the temporal modulation, the aSSR measurement becomes very noisy and unreliable.

Phase coherence across channels is a useful tool to investigate synchronization between distinct neuron populations. It can be generalized to multiple spectral components of neural responses and brain areas. It can measure coherence either within stimulus-induced oscillations, such as alpha and beta oscillations, or coherence between stimulus-induced and stimulus-evoked oscillations (such as aSSR). It can also address the relationship either within sensory cortex or between sensory cortex and the PFC. In addition, the direction of information flow can be determined to further evaluate the causality between the two correlated brain areas (Brancucci et al., 2005). Note that the channel pair should be separated by at least 100 mm to measure coherence between distinct sources of neural activity rather than the same sources (Srinivasan, Russell, Edelman, & Tononi, 1999).

We also explore blind source separation methods by independent component analysis (ICA), which decompose MEG responses into the sum of spatially fixed and distinct, temporally maximally independent component processes. ICA provides a new and potentially useful way to connect macroscopic activities with source activities. In the future, ICA decomposition might be combined with source localization algorithms to improve the convergence of the algorithm. Further studies will be required to fully assess the value and limitations of temporal independence as a segregation criterion. Interpretations of independent components from a neural point of view are also a direction of research.

5.2.2 Research on attention

Attention is one of the most intensely studied topics within psychology and cognitive neuroscience. The concept of attention was first proposed by William James. In his monumental Principles of Psychology (1890), he remarked:

> *"Everyone knows what attention is. It is the taking possession by the mind, in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought. Focalization, concentration, of consciousness are of its essence. It implies withdrawal from some things in order to deal effectively with others, and is a condition which has a real opposite in the confused, dazed, scatterbrained state which in French is called distraction, and Zerstreutheit in German."*

In James' time, the method mainly used to study attention was introspection, defined as self-observation and reporting of conscious inner thoughts, desires, and sensations. One major debate in that period was whether it was possible to attend to two things at once. In the 1950s, attention was first admitted to be a legitimate object of scientific study. At that time, the major debate was between early-selection models, which state that attention shuts down processing at the periphery level before the central level can analyze its content (Broadbent, 1958), and late-selection models, which postulate that the content is analyzed at the central level, but the unattended object cannot access consciousness (Deutsch, Deutsch, 1963).

The study of attention from a neuroscience perspective has been a challenging task. Approximately 50 years after the publication of David Hubel’s finding (1959) on the ‘attention units’ in the auditory cortex, it is still hotly debated what the proper subject for the empirical investigation of attention should be. Should attention be thought of as a vague and powerful ghost who influences our daily life but has no supporting substance? Or should attention be concreted as a computation or interaction between a separate set of neural areas with domain-specific systems? New techniques and methodological developments have opened the way for the prosperous study of cognitive neuroscience. Ambitious scientists are trying to quantify and localize high-level functions, such as attention and working memory. Nevertheless, because of the complex and highly dynamic nature of the brain, it is still a long and winding road before we can truly understand our own brain.

BIBIIOGRAPHY

Ahmar, N. E., & Simon, J. Z. (2005). MEG Adaptive Noise Suppression using Fast LMS. *Neural Engineering, 2005. Conference Proceedings. 2nd International IEEE EMBS Conference on*, 29-32.

Ahmar, N. E., Wang, Y., & Simon, J. Z. (2005). Significance tests for MEG response detection. *Neural Engineering, 2005. Conference Proceedings. 2nd International IEEE EMBS Conference on*, 21-24.

Ahveninen, J., Jaaskelainen, I. P., Raij, T., Bonmassar, G., Devore, S., Hamalainen, M., et al. (2006). Task-modulated" what" and" where" pathways in human auditory cortex. *Proceedings of the National Academy of Sciences, 103*(39), 14608.

Alcaini, M., Giard, M. H., Echallier, J. F., & Pernier, J. (1995). Selective auditory attention effects in tonotopically organized cortical areas: A topographic ERP study. *Human Brain Mapping, 2*(1), 59-169.

Anstis, S., & Saida, S. (1985). Adaptation to auditory streaming of frequency-modulated tones. *Journal of experimental psychology. Human perception and performance, 11*(3), 257-271.

Bacon, S. P., & Viemeister, N. F. (1985). Temporal modulation transfer functions in normal-hearing and hearing-impaired listeners. *Audiology, 24*(2), 117-134.

Baillet, S., & Garnero, L. (1997). A Bayesian approach to introducing anatomo-functional priors in theEEG/MEG inverse problem. *Biomedical Engineering, IEEE Transactions on, 44*(5), 374-385.

Baillet, S., Mosher, J., Leahy R.M. (2001). Electromagnetic Brain Mapping, *IEEE Signal Processing, Engineering, IEEE Transactions on, 18*, 14-30.

Bee, M. A., & Klump, G. M. (2005). Auditory Stream Segregation in the Songbird Forebrain: Effects of Time Intervals on Responses to Interleaved Tone Sequences. *Brain Behav Evol, 66*(3), 197-214.

Benda, J., Longtin, A., & Maler, L. (2006). A Synchronization-Desynchronization Code for Natural Communication Signals. *Neuron, 52*(2), 347-358.

Bendat, J. S., & Piersol, A. G. (1986). Random Data: Analysis and Measurement. *New York: John Wiley, 142*, 15.

Bendor, D., & Wang, X. (2005). The neuronal representation of pitch in primate auditory cortex. *Nature, 436*(7054), 1161-1165.

Bidet-Caulet, A., Fischer, C., Besle, J., Aguera, P. E., Giard, M. H., & Bertrand, O. (2007). Effects of Selective Attention on the Electrophysiological Representation of Concurrent Sounds in the Human Auditory Cortex. *Journal of Neuroscience, 27*(35), 9252.

Bieser, A., & Muller-Preuss, P. (1996). Auditory responsive cortex in the squirrel monkey: neural responses to amplitude-modulated sounds. *Experimental Brain Research, 108*(2), 273-284.

Bilecen, D., Seifritz, E., Scheffler, K., Henning, J., & Schulte, A. C. (2002). Amplitopicity of the Human Auditory Cortex: An fMRI Study. *Neuroimage, 17*(2), 710-718.

Bodnar, D. A., & Bass, A. H. (1997). Temporal Coding of Concurrent Acoustic Signals in Auditory Midbrain. *Journal of Neuroscience, 17*(19), 7553.

Brancucci, A., Babiloni, C., Vecchio, F., Galderisi, S., Mucci, A., Tecchio, F., et al. (2005). Decrease of functional coupling between left and right auditory cortices during dichotic listening: An electroencephalography study. *Neuroscience, 136*(1), 323-332.

Bregman, A. S. (1978). Auditory streaming is cumulative. *J Exp Psychol Hum Percept Perform, 4*(3), 380-387.

Bregman, A. S. (1990). *Auditory scene analysis*: MIT Press Cambridge, Mass.

Bregman, A. S., & Campbell, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. *J Exp Psychol, 89*(2), 244-249.

Broadbent, D. E. (1958). Perception and communication, New York, Pergamon Press.

Budd, T. W., Barry, R. J., Gordon, E., Rennie, C., & Michie, P. T. (1998). Decrement of the N1 auditory event-related potential with stimulus repetition: habituation vs. refractoriness. *International Journal of Psychophysiology, 31*(1), 51-68.

Buschman, T. J., & Miller, E. K. (2007). Top-Down Versus Bottom-Up Control of Attention in the Prefrontal and Posterior Parietal Cortices. *Science, 315*(5820), 1860.

Campbell, F. W., Atkinson, J., Francis, M. R., & Green, D. M. (1977). Estimation of auditory threshold using evoked potentials. *Auditory Evoked Potentials in Man. Psychopharmacology Correlates of EPs. Prog Clin Neurophysiol. Basel, Karger, 2*, 68?68.

Carlyon, R. P. (2004). How the brain separates sounds. *Trends in Cognitive Sciences, 8*(10), 465-471.

Carlyon, R. P., Cusack, R., Foxton, J. M., & Robertson, I. H. (2001). Effects of Attention and Unilateral Neglect on Auditory Stream Segregation. *Journal of Experimental Psychology, 27*(1), 115-127.

Cherry, E. C. (1953). Some Experiments on the Recognition of Speech, with One and with Two Ears. *The Journal of the Acoustical Society of America, 25*, 975.

Coch, D., Sanders, L. D., & Neville, H. J. (2005). An Event-related Potential Study of Selective Auditory Attention in Children and Adults. *Journal of Cognitive Neuroscience, 17*(4), 605-622.

Crick, F., & Koch, C. (1995). Are we aware of neural activity in primary visual cortex? *Nature, 375*(6527), 121-123.

Cusack, R., & Roberts, B. (2000). Effects of differences in timbre on sequential grouping. *Percept. Psychophys, 62*, 1112?1120.

da Silva, F. H. L. (1991). Neural mechanisms underlying brain waves: from neural membranes to networks. *Electroencephalogr. Clin. Neurophysiol, 79*(2), 81?83.

Dau, T., Kollmeier, B., & Kohlrausch, A. (1997). Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration. *The Journal of the Acoustical Society of America, 102*, 2906.

deCharms, R. C., Blake, D. T., & Merzenich, M. M. (1998). Optimizing Sound Features for Cortical Neurons. *Science, 280*(5368), 1439.

Desimone, R., & Duncan, J. (1995). Neural Mechanisms of Selective Visual Attention. *Annual Review of Neuroscience, 18*(1), 193-222.

Deustcsh, J. A. & Deustcsh, D. (1963). Attention: some theoretical considerations, *Psychological Review*, *70*, 80-90.

Dolphin, W. F. (1996). Auditory evoked responses to amplitude modulated stimuli consisting of multiple envelope components. *Journal of Comparative Physiology A: Sensory, Neural, and Behavioral Physiology, 179*(1), 113-121.

Dolphin, W. F. (1997). The envelope following response to multiple tone pair stimuli. *Hearing Research, 110*(1-2), 1-14.

Draganova, R., Ross, B., Borgmann, C., & Pantev, C. (2002). Auditory cortical response patterns to multiple rhythms of AM sound. *Ear Hear, 23*(3), 254-265.

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

Duncan, J., Emslie, H., Williams, P., Johnson, R., & Freer, C. (1996). Intelligence and the Frontal Lobe: The Organization of Goal-Directed Behavior. *Cognitive Psychology, 30*(3), 257-303.

Durif, C., Jouffrais, C., & Rouiller, E. M. (2003). Single-unit responses in the auditory cortex of monkeys performing a conditional acousticomotor task. *Experimental Brain Research, 153*(4), 614-627.

Efron, B., & Tibshirani, R. (1993). *An Introduction to the Bootstrap*. New York: Chapman & Hall/CRC.

Eggermont, J. J. (1998). Representation of Spectral and Temporal Sound Features in Three Cortical Fields of the Cat. Similarities Outweigh Differences. *Journal of Neurophysiology, 80*(5), 2743-2764.

Eggermont, J. J. (1999). The Magnitude and Phase of Temporal Modulation Transfer Functions in Cat Auditory Cortex. *Journal of Neuroscience, 19*(7), 2780.

Elhilali, M., Fritz, J. B., Klein, D. J., Simon, J. Z., & Shamma, S. A. (2004). Dynamics of Precise Spike Timing in Primary Auditory Cortex. *Journal of Neuroscience, 24*(5), 1159.

Eliades, S. J., & Wang, X. (2005). Dynamics of Auditory-Vocal Interaction in Monkey Auditory Cortex. *Cerebral Cortex, 15*(10), 1510-1523.

Fisher, N. I. (1993). *Statistical Analysis of Circular Data*. New York: Cambridge University Press.

Fishman, Y. I., Reser, D. H., Arezzo, J. C., & Steinschneider, M. (2001). Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey. *Hearing Research, 151*(1-2), 167-187.

Fishman, Y. I., & Steinschneider, M. (2006). Spectral Resolution of Monkey Primary Auditory Cortex (A1) Revealed With Two-Noise Masking. *Journal of Neurophysiology, 96*(3), 1105.

Fritz, J., Shamma, S., Elhilali, M., & Klein, D. (2003). Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nature Neuroscience, 6*(11), 1216-1223.

Fritz, J. B., Elhilali, M., David, S. V., & Shamma, S. A. (2007). Auditory attention -- ocusing the searchlight on sound. *Current Opinion in Neurobiology, 17*(4), 437-455.

Fritz, J. B., Elhilali, M., & Shamma, S. A. (2005). Differential Dynamic Plasticity of A1 Receptive Fields during Multiple Spectral Tasks. *Journal of Neuroscience, 25*(33), 7623.

Fujiwara, N., Nagamine, T., Imai, M., Tanaka, T., & Shibasaki, H. (1998). Role of the primary auditory cortex in auditory selective attention studied by whole-head neuromagnetometer. *Cognitive Brain Research, 7*(2), 99-109.

Gaese, B. H., & Ostwald, J. (1995). Temporal coding of amplitude and frequency modulation in the rat auditory cortex. *Eur J Neurosci, 7*(3), 438-450.

Gaetz, M., Weinberg, H., Rzempoluck, E., & Jantzen, K. J. (1998). Neural network classifications and correlation analysis of EEG and MEG activity accompanying spontaneous reversals of the Necker cube. *Cognitive Brain Research, 6*(4), 335-346.

Galambos, R., Makeig, S., & Talmachoff, P. J. (1981). A 40-Hz Auditory Potential Recorded from the Human Scalp. *Proceedings of the National Academy of Sciences of the United States of America, 78*(4), 2643-2647.

Ghose, G. M., & Maunsell, J. H. R. (2002). Attentional modulation in visual cortex depends on task timing. *Nature, 419*(6907), 616-620.

Giard, M. H., Collet, L., Bouchet, P., & Pernier, P. (1994). Auditory selective attention in the human cochlea. *Brain research, 633*(1-2), 353-356.

Gilbert, C. D., & Sigman, M. (2007). Brain States: Top-Down Influences in Sensory Processing. *Neuron, 54*(5), 677-696.

Giraud, A. L., Lorenzi, C., Ashburner, J., Wable, J., Johnsrude, I., Frackowiak, R., et al. (2000). Representation of the Temporal Envelope of Sounds in the Human Brain. *Journal of Neurophysiology, 84*(3), 1588-1598.

Gottlieb, J. P., Kusunoki, M., & Goldberg, M. E. (1998). The representation of visual salience in monkey parietal cortex. *Nature, 391*, 481-484.

Greenberg, S., Carvey, H., Hitchcock, L., & Chang, S. (2003). Temporal properties of spontaneous speech 梐 syllable-centric perspective. *Journal of Phonetics, 31*(3-4), 465-485.

Griffiths, T. D., & Warren, J. D. (2004). What is an auditory object? *Nat Rev Neurosci, 5*(11), 887-892.

Grimault, N., Micheyl, C., Carlyon, R. P., & Collet, L. (2002). Evidence for two pitch encoding mechanisms using a selective auditory training paradigm. *Perception & Psychophysics, 64*(2), 189-197.

Gruber, W. R., Klimesch, W., Sauseng, P., & Doppelmayr, M. (2005). Alpha Phase Synchronization Predicts P1 and N1 Latency and Amplitude Size. *Cerebral Cortex, 15*(4), 371-377.

Gutschalk, A., Micheyl, C., Melcher, J. R., Rupp, A., Scherg, M., & Oxenham, A. J. (2005). Neuromagnetic Correlates of Streaming in Human Auditory Cortex. *Journal of Neuroscience, 25*(22), 5382.

Hall, J. W., & Harvey, A. D. G. (1984). NoSo and NoSp thresholds as a function of masker level for narrow-band and wideband masking noise. *The Journal of the Acoustical Society of America, 76*, 1699.

Hamalainen, M., Hari, R., Ilmoniemi, R. J., Knuutila, J., & Lounasmaa, O. V. (1993). Magnetoencephalography -- heory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of Modern Physics, 65*(2), 413-497.

Hari, R., Aittoniemi, K., Jarvinen, M. L., Katila, T., & Varpula, T. (1980). Auditory evoked transient and sustained magnetic fields of the human brain localization of neural generators. *Experimental Brain Research, 40*(2), 237-240.

Hari, R., Hamalainen, M., & Joutsiniemi, S. L. (1989). Neuromagnetic steady-state responses to auditory stimuli. *The Journal of the Acoustical Society of America, 86*, 1033.

Harms, M. P., & Melcher, J. R. (2002). Sound Repetition Rate in the Human Auditory Pathway: Representations in the Waveshape and Amplitude of fMRI Activation. *Journal of Neurophysiology, 88*(3), 1433-1450.

Hart, H. C., Palmer, A. R., & Hall, D. A. (2003). Amplitude and Frequency-modulated Stimuli Activate Common Regions of Human Auditory Cortex. *Cerebral Cortex, 13*(7), 773-781.

Hartmann, W. M., & Johnson, D. (1991). Stream segregation and peripheral channeling. *Music Perception, 9*(2), 155-184.

Haynes, J. D., & Rees, G. (2005). Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nature Neuroscience, 8*(5), 686-691.

Heil, P., & Irvine, D. R. F. (1998). Functional Specialization in Auditory Cortex: Responses to Frequency-Modulated Stimuli in the Cat's Posterior Auditory Field. *Journal of Neurophysiology, 79*(6), 3041-3059.

Hillyard, S. A. (1998). Sensory gain control (amplification) as a mechanism of selective attention: electrophysiological and neuroimaging evidence. *Philosophical Transactions: Biological Sciences, 353*(1373), 1257-1270.

Hillyard, S. A., Hink, R. F., Schwent, V. L., & Picton, T. W. (1973). Electrical Signs of Selective Attention in the Human Brain. *Science, 182*(4108), 177-180.

Hubel, D. H., Henson, C. O., Rupert, A., & Galambos, R. (1959). " Attention" Units in the Auditory Cortex. *Science, 129*(3358), 1279-1280.

ISO-226. (2003). Acoustics—Normal equal-loudness-level contours. *International Organization for Standardization, Geneva, Switzerland.*

Jaaskelainen, I. P., Ahveninen, J., Belliveau, J. W., Raij, T., & Sams, M. (2007). Short-term plasticity in auditory cognition. *Trends in Neurosciences, 30*(12), 653-661.

Jaaskelainen, I. P., Ahveninen, J., Bonmassar, G., Dale, A. M., Ilmoniemi, R. J., Levanen, S., et al. (2004). Human posterior auditory cortex gates novel sounds to consciousness. *Proceedings of the National Academy of Sciences, USA, 101*, 6809?6814.

Jancke, L., Mirzazade, S., & Joni Shah, N. (1999). Attention modulates activity in the primary and the secondary auditory cortex: a functional magnetic resonance imaging study in human subjects. *Neuroscience Letters, 266*(2), 125-128.

John, M. S., Lins, O. G., Boucher, B. L., & Picton, T. W. (1998). Multiple auditory steady-state responses (MASTER): stimulus and recording parameters. *Audiology, 37*(2), 59-82.

Joris, P. X., Schreiner, C. E., & Rees, A. (2004). Neural Processing of Amplitude-Modulated Sounds. *Physiological Reviews, 84*(2), 541-577.

Kauramaki, J., Jaaskelainen, I. P., & Sams, M. (2007). Selective Attention Increases Both Gain and Feature Selectivity of the Human Auditory Cortex. *PLoS ONE, 2*(9).

Kay, S. M. (1993). *Fundamentals of statistical signal processing: estimation theory*: Prentice-Hall, Inc. Upper Saddle River, NJ, USA.

Kayser, C., Petkov, C. I., Lippert, M., & Logothetis, N. K. (2005). Mechanisms for Allocating Auditory Attention: An Auditory Saliency Map. *Current Biology, 15*(21), 1943-1947.

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

Kilgard, M. P., & Merzenich, M. M. (1998a). Cortical Map Reorganization Enabled by Nucleus Basalis Activity. *Science, 279*(5357), 1714.

Kilgard, M. P., & Merzenich, M. M. (1998b). Plasticity of temporal information processing in the primary auditory cortex. *Nat Neurosci, 1*, 727-731.

Knudsen, E. I. (2007). Fundamental Components of Attention. *Annu Rev Neurosci, 30*, 57?58.

Kowalski, N., Depireux, D. A., & Shamma, S. A. (1996). Analysis of dynamic spectra in ferret primary auditory cortex. I. Characteristics of single-unit responses to moving ripple spectra. *Journal of Neurophysiology, 76*(5), 3503-3523.

Kuwada, S., Batra, R., & Maher, V. L. (1986). Scalp potentials of normal and hearing-impaired subjects in response to sinusoidally amplitude-modulated tones. *Hearing research, 21*(2), 179-192.

Lewicki, M. S. (2002). Efficient coding of natural sounds. *Nature Neuroscience, 5*(4), 356-363.

Li, W., Piech, V., & Gilbert, C. D. (2004). Perceptual learning and top-down influences in primary visual cortex. *Nature Neuroscience, 7*(6), 651-657.

Liang, L., Lu, T., & Wang, X. (2002). Neural Representations of Sinusoidal Amplitude and Frequency Modulations in the Primary Auditory Cortex of Awake Primates. *Journal of Neurophysiology, 87*(5), 2237-2261.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21*(1), 1-36.

Liegeois-Chauvel, C., Lorenzi, C., Trebuchon, A., Regis, J., & Chauvel, P. (2004). Temporal Envelope Processing in the Human Left and Right Auditory Cortices. *Cerebral Cortex, 14*(7), 731-740.

Lins, O. G., & Picton, T. W. (1995). Auditory steady-state responses to multiple simultaneous stimuli. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section, 96*(5), 420-432.

Lu, T., Liang, L., & Wang, X. (2001). Temporal and rate representations of time-varying signals in the auditory cortex of awake primates. *Nat Neurosci, 4*(11), 1131-1138.

Lukas, J. H. (1981). The Role of Efferent Inhibition in Human Auditory Attention: An Examination of the Auditory Brainstem Potentials. *International Journal of Neuroscience, 12*(2), 137-145.

Luo, H., & Poeppel, D. (2007). Phase Patterns of Neuronal Responses Reliably Discriminate Speech in Human Auditory Cortex. *Neuron, 54*(6), 1001-1010.

Lutkenhoner, B., Krumbholz, K., Lammertmann, C., Seither-Preisler, A., Steinstrater, O., & Patterson, R. D. (2003). Localization of Primary Auditory Cortex in Humans by Magnetoencephalography. *Neuroimage, 18*(1), 58-66.

Machens, C. K., Gollisch, T., Kolesnikova, O., & Herz, A. V. M. (2005). Testing the Efficiency of Sensory Coding with Optimal Stimulus Ensembles. *Neuron, 47*(3), 447-456.

Makeig, S., Westerfield, M., Jung, T. P., Enghoff, S., Townsend, J., Courchesne, E., et al. (2002). Dynamic Brain Sources of Visual Evoked Responses (Vol. 295, pp. 690-694).

Malone, B. J., Scott, B. H., & Semple, M. N. (2007). Dynamic Amplitude Coding in the Auditory Cortex of Awake Rhesus Macaques. *Journal of Neurophysiology, 98*(3), 1451.

Martinez-Trujillo, J. C., & Treue, S. (2002). Attentional Modulation Strength in Cortical Area MT Depends on Stimulus Contrast. *Neuron, 35*(2), 365-370.

Maunsell, J. H. R., & Treue, S. (2006). Feature-based attention in visual cortex. *Trends in Neurosciences, 29*(6), 317-322.

Micheyl, C., Tian, B., Carlyon, R. P., & Rauschecker, J. P. (2005). Perceptual Organization of Tone Sequences in the Auditory Cortex of Awake Macaques. *Neuron, 48*(1), 139-148.

Miller, L. M., Escabi, M. A., Read, H. L., & Schreiner, C. E. (2002). Spectrotemporal Receptive Fields in the Lemniscal Auditory Thalamus and Cortex. *Journal of Neurophysiology, 87*(1), 516-527.

Moore, B. C. J., & Gockel, H. (2002). Factors influencing sequential stream segregation. *Acta Acustica United with Acustica, 88*, 320-333.

Moore, D. R. (2003). Cortical neurons signal sound novelty. *Nature Neuroscience, 6*(4), 330-332.

Naatanen, R. (1990). The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. *Behavioral and Brain Sciences, 13*(2), 201-288.

Naatanen, R. (1992). *Attention and Brain Function*: Erlbaum.

Naatanen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology, 24*(4), 375-425.

Naatanen, R., Tervaniemi, M., Sussman, E., Paavilainen, P., & Winkler, I. (2001). "Primitive intelligence" in the auditory cortex. *Trends in Neurosciences, 24*(5), 283-288.

Narayan, R., Grana, G., & Sen, K. (2006). Distinct Time Scales in Cortical Discrimination of Natural Sounds in Songbirds. *Journal of Neurophysiology, 96*(1), 252.

Neff, D. L., & Green, D. M. (1987). Masking produced by spectral uncertainty with multicomponent maskers. *Percept Psychophys, 41*(5), 409-415.

Nelken, I. (2004). Processing of complex stimuli and natural scenes in the auditory cortex. *Current Opinion in Neurobiology, 14*(4), 474-480.

Niebur, E., Hsiao, S. S., & Johnson, K. O. (2002). Synchrony: a neuronal mechanism for attentional selection? *Current Opinion in Neurobiology, 12*(2), 190-194.

Okamoto, H., Ross, B., Kakigi, R., Kubo, T., & Pantev, C. (2004). N1m recovery from decline after exposure to noise with strong spectral contrasts. *Hearing Research, 196*(1-2), 77-86.

Okamoto, H., Stracke, H., Wolters, C. H., Schmael, F., & Pantev, C. (2007). Attention Improves Population-Level Frequency Tuning in Human Auditory Cortex. *Journal of Neuroscience, 27*(39), 10383.

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia, 9*(1), 97-113.

Pantev, C., Roberts, L. E., Elbert, T., Ro, B., & Wienbruch, C. (1996). Tonotopic organization of the sources of human auditory steady-state responses. *Hearing Research, 101*(1-2), 62-74.

Penny, W. D., Kiebel, S. J., Kilner, J. M., & Rugg, M. D. (2002). Event-related brain dynamics. *Trends in Neurosciences, 25*(8), 387-389.

Picton, T. W., John, M. S., Dimitrijevic, A., & Purcell, D. (2003). Human auditory steady-state responses. *Int J Audiol, 42*(4), 177-219.

Picton, T. W., Woods, D. L., & Proulx, G. B. (1978a). Human auditory sustained potentials. I. The nature of the response. *Electroencephalogr Clin Neurophysiol, 45*(2), 186-197.

Picton, T. W., Woods, D. L., & Proulx, G. B. (1978b). Human auditory sustained potentials. II. Stimulus relationships. *Electroencephalogr Clin Neurophysiol, 45*(2), 198-210.

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as 詁 symmetric sampling in time? *Speech Communication, 41*, 245?255.

Poeppel, D., Yellin, E., Phillips, C., Roberts, T. P. L., Rowley, H. A., Wexler, K., et al. (1996). Task-induced asymmetry of the auditory evoked M100 neuromagnetic field elicited by speech sounds. *Cognitive Brain Research, 4*(4), 231-242.

Polley, D. B., Steinberg, E. E., & Merzenich, M. M. (2006). Perceptual Learning Directs Auditory Cortical Map Reorganization through Top-Down Influences. *Journal of Neuroscience, 26*(18), 4970.

Posner, M. I., & Petersen, S. E. (1990). The Attention System of the Human Brain. *Annual Review of Neuroscience, 13*(1), 25-42.

Pressnitzer, D., & Hup, J. M. (2005). Is auditory streaming a bistable percept? *Proceedings of Forum Acusticum, Budapest*, 1557-1561.

Qiu, F. T., Sugihara, T., & von der Heydt, R. (2007). Figure-ground mechanisms provide structure for selective attention. *Nat Neurosci*.

Rees, A., Green, G. G., & Kay, R. H. (1986). Steady-state evoked responses to sinusoidally amplitude-modulated sounds recorded in man. *Hear Res, 23*(2), 123-133.

Regan, D. (1966). Some characteristics of average steady-state and transient responses evoked by modulated light. *Electroencephalogr Clin Neurophysiol, 20*(3), 238-248.

Regan, D., & Heron, J. R. (1969). Clinical investigation of lesions of the visual pathway: a new objective technique. *J Neurol Neurosurg Psychiatry, 32*(5), 479-483.

Regan, M. P., & Regan, D. (1988). A frequency domain technique for characterizing nonlinearities in biological systems. *Journal of theoretical biology, 133*(3), 293-317.

Reynolds, J. H., & Chelazzi, L. (2004). ATTENTIONAL MODULATION OF VISUAL PROCESSING. *Annual Review of Neuroscience, 27*(1), 611-647.

Reynolds, J. H., Chelazzi, L., & Desimone, R. (1999). Competitive Mechanisms Subserve Attention in Macaque Areas V2 and V4. *Journal of Neuroscience, 19*(5), 1736.

Roberts, T. P. L., Ferrari, P., Stufflebeam, S. M., & Poeppel, D. (2000). Latency of the auditory evoked neuromagnetic field components: stimulus dependence and insights toward perception. *J Clin Neurophysiol, 17*(2), 114-129.

Roelfsema, P. R., Tolboom, M., & Khayat, P. S. (2007). Different Processing Phases for Features, Figures, and Selective Attention in the Primary Visual Cortex. *Neuron, 56*(5), 785-792.

Ross, B., Borgmann, C., Draganova, R., Roberts, L. E., & Pantev, C. (2000). A high-precision magnetoencephalographic study of human auditory steady-state responses to amplitude-modulated tones. *The Journal of the Acoustical Society of America, 108*, 679.

Saalmann, Y. B., Pigarev, I. N., & Vidyasagar, T. R. (2007). Neural Mechanisms of Visual Attention: How Top-Down Feedback Highlights Relevant Locations. *Science, 316*(5831), 1612.

Schreiner, C. E., & Urbas, J. V. (1988). Representation of amplitude modulation in the auditory cortex of the cat. II. Comparison between cortical fields. *Hear Res, 32*(1), 49-63.

Schulze, H., & Langner, G. (1997). Periodicity coding in the primary auditory cortex of the Mongolian gerbil (Merionesunguiculatus?): two different coding strategies for pitch and rhythm? *Journal of Comparative Physiology A: Sensory, Neural, and Behavioral Physiology, 181*(6), 651-663.

Shahin, A., Bosnyak, D. J., Trainor, L. J., & Roberts, L. E. (2003). Enhancement of Neuroplastic P2 and N1c Auditory Evoked Potentials in Musicians. *Journal of Neuroscience, 23*(13), 5545.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech Recognition with Primarily Temporal Cues. *Science, 270*(5234), 303.

Simon, J. Z., & Wang, Y. (2005). Fully complex magnetoencephalography. *Journal of Neuroscience Methods, 149*(1), 64-73.

Singer, W., & Strategies, T. C. B. (1999). Neuronal Synchrony: A Versatile Code Review for the Definition of Relations? *Neuron, 24*, 49-65.

Smith, E. C., & Lewicki, M. S. (2006). Efficient auditory coding. *Nature, 439*(7079), 978-982.

Snyder, J. S., Alain, C., & Picton, T. W. (2006). Effects of Attention on Neuroelectric Correlates of Auditory Stream Segregation. *Journal of Cognitive Neuroscience, 18*(1), 1-13.

Srinivasan, R., Russell, D. P., Edelman, G. M., & Tononi, G. (1999). Increased Synchronization of Neuromagnetic Responses during Conscious Perception. *Journal of Neuroscience, 19*(13), 5435.

Sussman, E., & Steinschneider, M. (2006). Neurophysiological evidence for context-dependent encoding of sensory input in human auditory cortex. *Brain Research, 1075*(1), 165-174.

Talavage, T. M., Sereno, M. I., Melcher, J. R., Ledden, P. J., Rosen, B. R., & Dale, A. M. (2004). Tonotopic Organization in Human Auditory Cortex Revealed by Progressions of Frequency Sensitivity. *Journal of Neurophysiology, 91*(3), 1282-1296.

Tallon-Baudry, C., & Bertrand, O. (1999). Oscillatory gamma activity in humans and its role in object representation. *Trends in Cognitive Sciences, 3*(4), 151-162.

Tiitinen, H. T., Sinkkonen, J., Reinikainen, K., Alho, K., Lavikainen, J., & Naeaetaenen, R. (1993). Selective attention enhances the auditory 40-Hz transient response in humans. *Nature, 364*(6432), 59-60.

Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*(1), 97-136.

Ulanovsky, N., Las, L., Farkas, D., & Nelken, I. (2004). Multiple Time Scales of Adaptation in Auditory Cortex Neurons. *Journal of Neuroscience, 24*(46), 10440.

Ulanovsky, N., Las, L., & Nelken, I. (2004). Processing of low-probability sounds by cortical neurons. *Neuronal Adaptation in Cat Auditory Cortex, 6*(4), 391-398.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. *Analysis of visual behavior*, 549-586.

van Noorden, L. (1975). Temporal Coherence in the Perception of Tone Sequences.

van Noorden, L. (1977). Minimum differences of level and frequency for perceptual fission of tone sequences ABAB. *The Journal of the Acoustical Society of America, 61*, 1041.

Vanni, S., Revonsuo, A., & Hari, R. (1997). Modulation of the Parieto-Occipital Alpha Rhythm during Object Detection. *Journal of Neuroscience, 17*(18), 7141.

Viemeister, N. F. (1979). Temporal modulation transfer functions based upon modulation thresholds. *The Journal of the Acoustical Society of America, 66*, 1364.

Viemeister, N. F., & Plack, C. J. (1993). Time analysis. *Human psychophysics, 3*, 116-154.

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

Wang, X., Lu, T., & Liang, L. (2003). Cortical processing of temporal modulations. *Speech Commun, 41*, 107-121.

Wang, X., Lu, T., Snider, R. K., & Liang, L. (2005). Sustained firing in auditory cortex evoked by preferred stimuli. *Nature, 435*(7040), 341-346.

Winkler, I., Karmos, G., & Naatanen, R. (1996). Adaptive modeling of the unattended acoustic environment reflected in the mismatch negativity event-related potential. *Brain Research, 742*(1-2), 239-252.

Wohlgemuth, S., & Ronacher, B. (2007). Auditory Discrimination of Amplitude Modulations Based on Metric Distances of Spike Trains. *Journal of Neurophysiology, 97*(4), 3082.

Woldorff, M. G., Gallen, C. C., Hampson, S. A., Hillyard, S. A., Pantev, C., Sobel, D., et al. (1993). Modulation of Early Sensory Processing in Human Auditory Cortex During Auditory Selective Attention. *Proceedings of the National Academy of Sciences of the United States of America, 90*(18), 8722-8726.

Woldorff, M. G., & Hillyard, S. A. (1991). Modulation of early auditory processing during selective listening to rapidly presented tones. *Electroencephalogr Clin Neurophysiol, 79*(3), 170-191.

Woolley, S. M., Fremouw, T. E., Hsu, A., & Theunissen, F. E. (2005). Tuning for spectro-temporal modulations as a mechanism for auditory discrimination of natural sounds. *Nat Neurosci, 8*(10), 1371-1379.

Xiang, J., Wang, Y., & Simon, J. Z. (2005). MEG Responses to Speech and Stimuli With Speechlike Modulations. *International IEEE EMBS Conference on Neural Engineering*.

Zani, A., & Proverbio, A. M. (1995). ERP signs of early selective attention effects to check size. *Electroencephalography and Clinical Neurophysiology, 95*(4), 277-292.

Zatorre, R. J., & Belin, P. (2001). Spectral and Temporal Processing in Human Auditory Cortex. *Cerebral Cortex, 11*(10), 946-953.