ABSTRACT

Title of Document: Tracking Sound Dynamics in Human Auditory Cortex: New macroscopic perspectives from MEG

Huan Luo, PhD, 2007

Directed By: Dr. David Poeppel
Professor
Neuroscience and Cognitive Science Program
Department of Biology
Department of Linguistics

Both the external world and our internal world are full of changing activities , and the question of how these two dynamic systems are linked constitutes the most intriguing and fundamental question in neuroscience and cognitive science. This study specifically investigates the processing and representation of sound dynamic information in human auditory cortex using magnetoencephalography (MEG), a non-invasive brain imaging technique whose high temporal resolution (on the order of ~1ms) makes it an appropriate tool for studying the neural correlates of dynamic auditory information.

The other goal of this study is to understand the essence of the macroscopic activities reflected in non-invasive brain imaging experiments, specifically focusing on MEG. Invasive single-cell recordings in animals have yielded a large amount of information about how the brain works at a microscopic level. However, there still

exist large gaps in our understanding of the relationship between the activities recorded at the microscopic level in animals and at the macroscopic level in humans, which have yet to be reconciled in terms of their different spatial scales and activities format, making a unified knowledge framework still unsuccessful.

In this study, natural speech sentences and sounds containing speech-like temporal dynamic features are employed to probe the human auditory system. The recorded MEG signal is found to be well correlated with the stimulus dynamics via amplitude modulation (AM) and/or phase modulation (PM) mechanisms. Specifically, oscillations at various frequency bands are found to be the main information-carrying elements of the MEG signal, and the two major parameters of these endogenous brain rhythms, amplitude and phase, are modulated by incoming sensory stimulus dynamics, corresponding to AM and PM mechanism, to track sound dynamics. Crucially, such modulation tracking is found to be correlated with human perception and behavior.

This study suggests that these two dynamic and complex systems, the external and internal worlds, systematically communicate and are coupled via modulation mechanism, leading to a reverberating flow of information embedded in oscillating waves in human cortex. The results also have implications for brain imaging studies, suggesting that these recorded macroscopic activities reflect ‘brain state’, the more close neural correlate of high-level cognitive behavior.

TRACKING SOUND DYNAMICS IN HUMAN AUDITORY CORTEX: NEW MACROSCOPIC PERSPECTIVES FROM MEG

By

Huan Luo

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
2007

Advisory Committee:
Professor David Poeppel, Chair
Professor Jonathan Z. Simon
Professor Barry Horwitz
Professor Todd Troyer
Professor Rodolfo R. Llinás
Professor Robert Dooling

Dedication

I dedicate this dissertation to my lovely family,

To my husband Zheng

to my two sons (F.M. and P.M.)

to my Mom and Dad

Without you, I don't know what I can do, and where I will go…

Acknowledgements

In retrospect of the 5 PhD years spent in Maryland, I feel really lucky to have met lots of nice and considerate persons who gave me so much help in both life and academics.

I am so lucky to have you, David, as my advisor, who convinced me to go to Maryland. I could still vividly recall the first time we met, at the airport 5 years ago, when I and my husband, two exciting but nervous young Chinese students, saw you, who picked us up, showed us around campus in beautiful sunset, and made us out of the homesick mode. Your tenet for graduate student is to give us freedom and generous support, and push us to actively pursue our own interest rather than to passively follow instructions, even that are not your own interests. It is this tenet that guides me to be an independent researcher, which will be my pursued goal in my lifetime career. I could not imagine that without your help and considerations, how could I gradually turn from a naïve and always-confused layman to a passionate researcher? How could I spend most of the year in China accompanying my son and pursue my degree at the same time and still got your financial support? How could I achieve the balance between work and family responsibility? Your attitude influences me a lot, not only in academics, but also in the view of cherishing my family. You are right that family is first-ranked in one's life and you are right that to investigate what attracts me most is the only worthwhile thing in academics.

I am luck to have you, Barry, as my co-advisor, who selected my personal statement from lots of applications just because you thought that my statement was frank and full of grammar mistakes. What an interesting reason and what a wonderful

experience! I really enjoy working with you, talking with you, arguing with you. Your patience and tolerance give me, a naïve student, so much courage to pursue in my own way, and your considerations give me freedom to spend my last 3 PhD years exploring my own interests. Thank you, Barry!

I am lucky to work with you, Jonathan, who gave me definitely the most solid academic help in my work. I am always surprised and impressed by your carefulness, speedy response, clarity, strong mathematics and sagacity, and I hope that one day I could give my students as much guidance as you gave to me. You are also the person, who was always doubtful and skeptical when I excitedly told you seemingly excellent findings, pushing me to dig hard and pursue in a more object and comprehensive way. I enjoyed so much the time when we sit together and examined the noisy data in the laptop, trying to decode information from them, and it was really a wonderful discovery journey!

I am lucky to meet and learn from you, Todd, who taught the 'computational neuroscience' course in my second year and I really learned a lot and got so much inspiration from this course. You introduced lots of intriguing computational methods employed in neurophysiology studies and lead me into a world with more order and possible ways to see, to hypothesize, and to summarize. From them, I began to think that I should jump from the traditional perspectives regarding MEG activities and borrowed more quantitative methods and ideas from the mature animal work, to pursue the order and precision in such a macroscopic world as found in microscopic world. I could also see your strong support for my work. Thank you, Todd!

I am also so lucky to have the chance to meet and talk with you, Dr. Llinas! It is really an unforgettable and unusual interview experience and I am so grateful to your offer. It was the most difficult presentation I did where you raised so much different questions and interpretations I never thought about. On the other hand, I got the most valuable suggestions and implications from it, and such an experience really opened me a new door to look at the macroscopic activities. As you said, you wanted me to think from a distinct perspective, which I am trying to do and become more and more convinced now. Although I am not sure it is completely right, I like such a new goal and for me, it is the beginning of new brain exercise and a new exciting pathway! I also appreciate your considerations about my family issues and your respect for my short-term plans, and I am sure that I will learn more from you in future.

I am lucky to be a member of NACS program and CNL lab. I found that people around are all so nice and helpful. Sandy, I do not know how to express my sincere appreciations for your help in so many detailed things. Avis, from you, I know how a woman scientist can achieve excellent balance between academic work and family. Richard, I still remembered the back-and-forth emails we had before I came to NACS, where you patiently answered all of my confusions and unease. Jeff, you are really really nice to help me a lot during MEG experiment recordings. I have a bad memory and poor set-up ability, and it was you who smoothly helped me overcome those moments. Susannah, thanks for your hard work in proofreading my manuscripts which are always very long and full of complex technique and grammar errors. Robert and Kathi, thanks for your help in so many detailed procedure things. Maria, my closest foreign friend and colleague, we come to Maryland at the same time, both

of us left our country, and are layman of this field coming here to challenge ourselves in this intriguing and new field. I learned a lot from you, your diligence, your smartness, your insistence, and your courage, and I am sure you will have a wonderful future!

I am also very lucky to have so many Chinese friends here. Haiyan and Feng, we have experienced so many things together in the 5 years, and we are like to live and grow up together, and we both have our kids now. How wonderful life is! Youjun & Xiaosong, Ping&Yirong, I really want to say thank you very much for your sincere help to my family in the 5 years.

Most luckily, I have a wonderful family! Mom, thank you for your help me in taking care of Maomao. I am not aware how great a mom is and how great you are until I become a mom. Your courage, your selflessness, your hardworking are things I need to learn in the long run. Dad, thank you for supporting me to do anything I like. I may do something which you never understand, but from your proud smile, I know you respect my decision and that is the drive for my life pathway ahead. Zheng, my dearest husband, we have been together for more than 10 years now and we have shared so many moments together, in smile, in tears, in depression and in excitement. Without you, I will not be what I am now, and I will not be a happy mom/wife and a passionate researcher. Without your respect and encouragement, I will not have so much courage to go on in this academic field. Maomao (Franklin Mutian), Mom owns you so much and could not accompany you as long as other parents did. I could see your love and smartness in your eye although you could not speak yet, and you are the cleverest kid in mom's eyes. My younger son, Kuokuo (Patrick Mukuo), mom

also want to thank you for accompanying me in the most difficult time, giving me so much inspirations and courage. I have not seen you yet, but could feel your energy and your love. You two are my biggest accomplishments and miracles in my life, and I am proud of you!

Table of Contents

List of Figures

Chapter 1: Introduction

We are living in a world full of changes in multiple dimensions. Even a stationary object changes in the spatial dimension, with different colors, textures and shapes occupying different spaces in the same object. In fact, it is just such a combination of 'change features' that constitutes the 'uniqueness' of this object. When the object is moving, its position in space changes continuously, resulting in changes in another important dimension—time. From a broader perspective, we encounter and expect changes all the time; life would become tedious and boring if the external world stayed the same every day.

The brain, a complex and important organ, is germane to our minds, thoughts, emotions, and importantly, the communication between our internal world and the complex, ever-changing outside world. Brains are made up of enormous numbers of neurons that work together to make us what we are. Interestingly, if we dig into the brain and look at its activities in real time, we will see again a rapidly changing world—individual neurons fire spikes occasionally even in the absence of any outside stimulus, neuron groups manifest changing activity, and local field potentials (LFP) display fluctuating patterns.

A central puzzle in neuroscience is how those individual, semiautonomous neurons in the brain work together to link our internal world with the outside world in

real time, to receive information from it and to act upon it. Specifically, how is the dynamic outside world embedded and represented in the (also dynamic) inside world? What do the richly dynamic brain activities we observe represent and reflect? Are they a simple reflection of changes in the outside world, or are they some form of abstraction from the complex external environment? Or are they not directly related to the external world? Do they perhaps represent the ever-changing internal 'mind states'?

This thesis explores the tracking and representation of auditory stimulus dynamics in human auditory cortex, and the results can extend to other sensory domains. The auditory domain is a very interesting topic to begin with, for two reasons. First, a sound, reflecting changes in acoustic pressure, is endowed with innate dynamic properties. It unfolds in time, and one needs to attend to the changes in real time to perceive the sound and extract the pertinent information it contains. Audition is a natural example that is well suited to the study of cortical mechanisms for representing stimulus feature changes. Secondly, cortical processing of speech, a unique and complex communicative capability only humans are endowed with, remains important but obscure in auditory neuroscience. Speech is made up of complex acoustic signals from the perspective of signal processing, containing rich dynamic structure in both amplitude and frequency. To study the neural correlates of speech recognition, the most useful and direct evidence is obtained using natural speech stimuli or stimuli with speech-like dynamics, and the results from studies employing simple pure tones or noise may not be directly applicable.

Experimental results depend on the approaches and tools employed, and different perspectives will lead to various interpretations. When testing the 'linking hypothesis' between the outside and inside worlds, confusions will arise if the assumptions, mechanisms and limitations of the specific experimental approach are not considered. In addition, to combine results from different experimental techniques, one needs to be aware of the differences and the relationships among groups of data, even when they are recorded from the same brain at the same time. Data recorded with different tools manifest distinct formats and patterns, and to grasp the subtle information buried in large amounts of noise, their respective essential features, or 'singular vectors', need to be estimated before any conclusion is made. For example, in single-cell recording in animals, the post-stimulus time histogram (PSTH), displaying the neuron's time-varying spike rate pattern, is a generally accepted quantitative way to describe the activity of single neurons, or microscopic activity. However, this becomes more complicated for data at larger spatial scales—neuron population activities, system activities, and human brain imaging data, including fMRI, EEG and MEG. This complexity is due to the indirect relationship between microscopic activity and macroscopic activity. As 'system theory' points out, it may be impossible to interpret the behavior of the whole in terms of the behavior of the parts.

The importance of dynamic properties in both natural sounds and brain activity will be detailed in section 1. In section 2, brain imaging techniques and various explanations of brain imaging data, especially the EEG/MEG data and their possible relationship with neurophysiological data, will be introduced. In sections 3 and 4,

previous results about the representation of sound dynamics in animals and humans, respectively, will be addressed, and a summary will be given in section 5.

1.1 Dynamics is crucial

1.1.1 What acoustic features characterize a sound?

A sound is a form of 'pressure wave' produced by a vibration or oscillation that causes a periodic disturbance of the surrounding air or other medium. A sound is audible to the human ear if its frequency (number of vibrations per second) falls between 20 Hz and 20,000 Hz. A direct illustration of a sound is a temporal waveform, a 2-D representation, which indicates the physical disturbance of a medium over time. The most useful and popular representation of a sound is the spectrogram, a 3-D plot of the conjunctional spectral and temporal energy as a function of time. A simple sound such as a pure tone can be uniquely described by its intensity, frequency, starting phase and duration, and its spectrogram manifests a horizontal line, indicating its stationary nature—its unchanging frequency. The spectrograms of more complex sounds are full of dynamic structures rather than a flat line, indicating that their frequency components change over time. Figure 1-1 shows the temporal waveforms and spectrograms of a simple pure tone and a human speech signal.

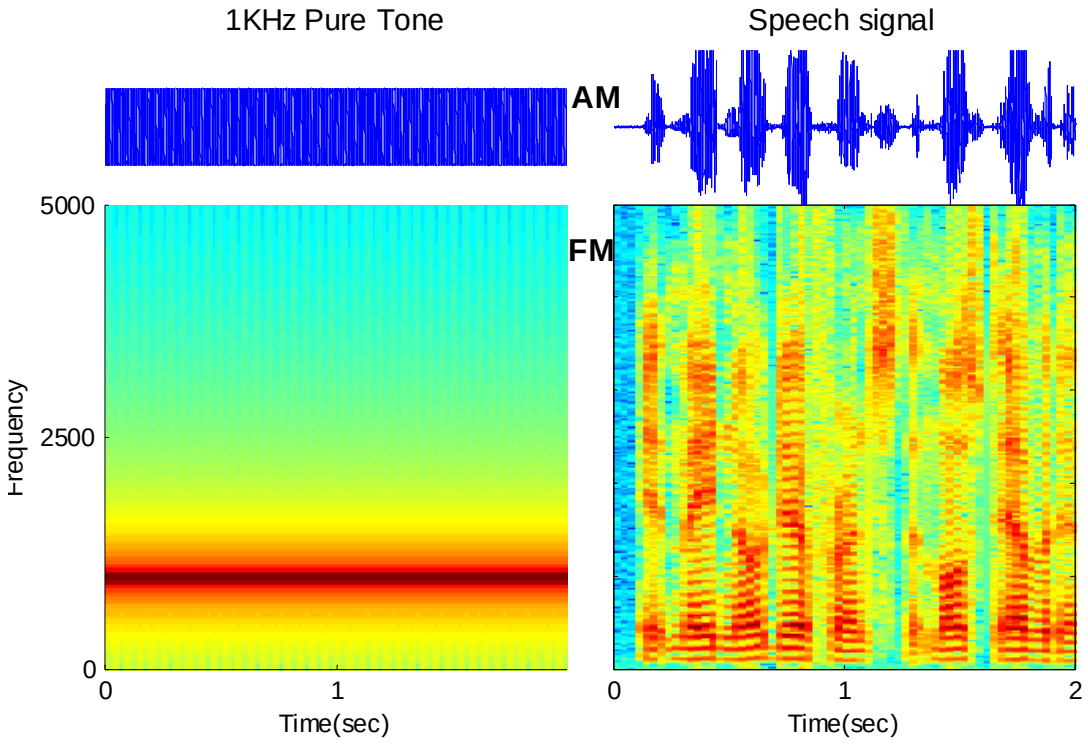


Figure 1-1 Simple sound example and complex sound example. Left: a 1 kHz pure tone. Right: a natural human speech signal. Upper panel: temporal waveform. Lower panel: spectrogram. Note that the complex speech signal contains rich dynamic structures in both amplitude (AM) and frequency (FM), compared to the simple pure tone.

Obviously, it is not enough to characterize complex sounds in terms of the 4 physical properties used to describe pure tones, since the properties (e.g., intensity and frequency here) change continuously as a function of time. A complex natural sound (e.g., human speech signal, etc.) is better characterized by a 3-D spectrogram, reflecting a specific spectrotemporal energy pattern. Note that this characteristic pattern contains components that change in both frequency (vertical axis) and amplitude (color) as a function of time (horizontal axis), defined as frequency

modulation (FM) and amplitude modulation (AM), respectively. In other words, to fully describe a complex sound, we need to introduce additional parameters or features to characterize its dynamic structures, including both FM and AM. One suitable way to extract and separate these two features is to perform a Hilbert transform of the sound, resulting in an envelope signal and a fine structure signal, each containing separate amplitude dynamic information and frequency dynamic information, corresponding to AM and FM features, respectively. Figure 1-2 illustrates the Hilbert transform of a complex signal.

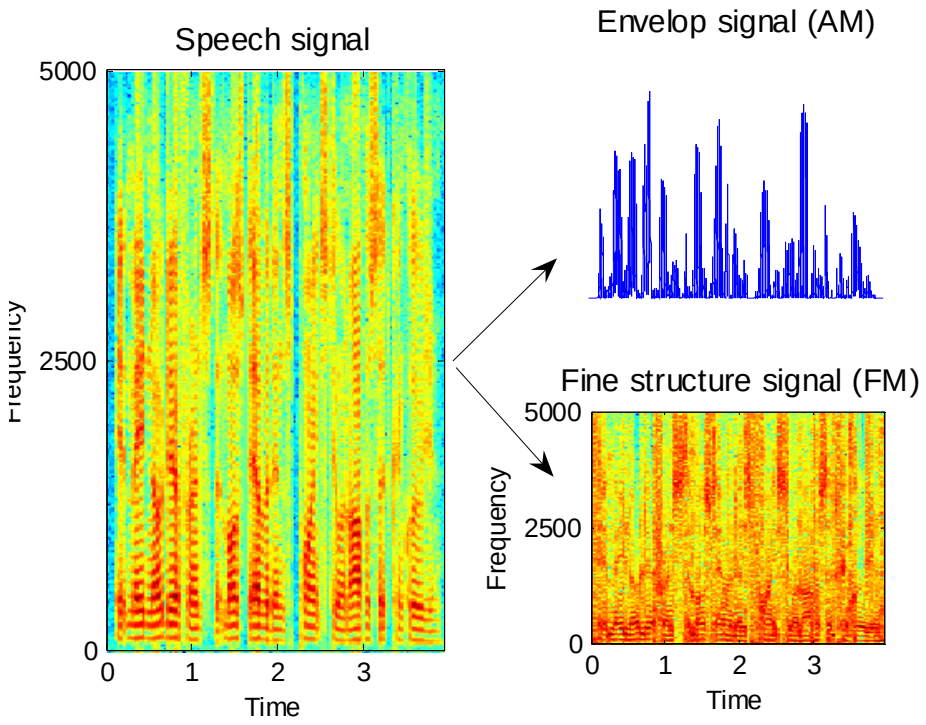


Figure 1-2 Hilbert transform of a speech signal. A complex signal is decomposed to an envelope signal and a fine structure signal, representing dynamics in amplitude (AM) and in frequency (FM).

Natural sounds, especially the species-specific communication sounds, contain rich dynamic structures in both amplitude and frequency. These acoustic transients occur within a broad variety of time scales, ranging from a few milliseconds to several hundred milliseconds and longer, and convey behaviorally relevant information. Thus, in contrast to relatively stationary simple pure tones, which can be characterized by stationary properties such as frequency content from performing a Fourier transform of the whole signal, the most necessary and important information about natural sounds lies in their temporal structures, depicted by corresponding AM and FM features.

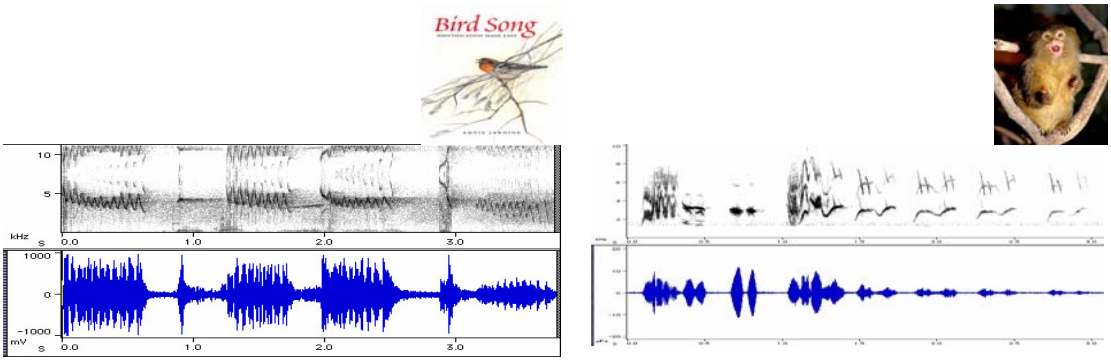


Figure 1-3 Species-specific communication sounds (Left: bird song; Right: marmoset calls) contain rich dynamic structures in both amplitude (AM, lower panel) and frequency (FM, upper panel).

1.1.2 What acoustic features do cortical neurons prefer?

Neurons in sensory cortices are often assumed to be 'feature detectors', computing simple and then successively more complex features from the incoming sensory stream (Nelken et al., 2004). This dominant assumption has evolved from success in understanding the visual cortex; it is therefore not difficult to understand how concept such as 'grandmother cell', a cell whose firing denotes the recognition or identification of one's grandmother, is coming from. Another dominant assumption in studying the sensory cortices is 'point-to-point' spatial mapping, which originates from the somatotopic mapping reported in earlier studies by Penfield and Rasmussen (1950).

These two assumptions also guide and dominate auditory neuroscience research. For example, a tonotopic map was found in auditory cortex; in other words, frequency, the most elementary property of sound, is represented in different places in the auditory cortex. This map was proposed to have originated from the tonotopic map in cochlear and subcortical structures along the auditory pathway, and it has become a routine paradigm to quantize the 'characteristic frequency' of each recorded auditory neuron. As to the species-specific vocalizations, large efforts are made to seek 'vocalization selective cells', analogous to the 'face cells' found in monkeys. There are other relatively stationary auditory properties, such as pitch, that have also been found to be encoded in clustered areas (Bendor and Wang, 2005).

As introduced in the previous section, temporal modulations are fundamental components of natural sounds and convey behaviorally important information. The

neural representations of temporal modulations are present throughout the auditory pathway, and at the auditory periphery, auditory nerve fibers discharge spike patterns that faithfully represent the temporal structure of sound. Such precise tracking and representation degrades as it goes to higher stages of processing, and by the time one reaches the auditory cortex, neurons can no longer follow rapidly changing stimuli. Note that the investigations of temporal modulation tracking of neurons actually examine their temporal resolutions instead of their possible role as 'temporal modulation feature detectors'. In other words, if temporal modulation is a feature that needs to be encoded and represented in auditory cortex, those auditory neurons should be selective for certain temporal modulation frequencies, or have a 'characteristic modulation frequency' in addition to their 'characteristic frequency'. Many findings suggest that temporal modulation features, especially the low frequency features, are widely represented in auditory cortex. The details of these findings will be addressed in section 3.

Compared to fixing the stimulus ensemble according to the acoustic properties being investigated, such as frequency, pitch, temporal modulation, etc., and testing auditory neurons by varying stimulus values along one dimension, a more objective approach to examining the preference properties of auditory neurons is to employ reverse correlation techniques or spike-triggered averaging techniques, using a more random and less controlled stimulus ensemble. It is a commonly exploited characteristic function of neurons that describes their 'response areas' or preference features. Using these techniques, it is shown that the response properties of auditory cortical neurons are dominated by transient changes in both amplitude and frequency,

reflecting their selectivity for AM and FM features in the stimuli (deCharms et al., 1998; Deprieux et al., 2001; Miller et al., 2002; Elhilali et al., 2004). Figure 1-4 gives several examples of the receptive fields of auditory cortical neurons. Interestingly, the reverse approach, which makes the theoretical assumption that the auditory system's encoding mechanisms are shaped to represent natural sounds in the most optimal and efficient way, predicts again a preponderance of AM and FM response patterns in the receptive fields of auditory cortical neurons (Lewicki, 2002; Klein et al., 2003), in agreement with empirical findings about receptive fields.

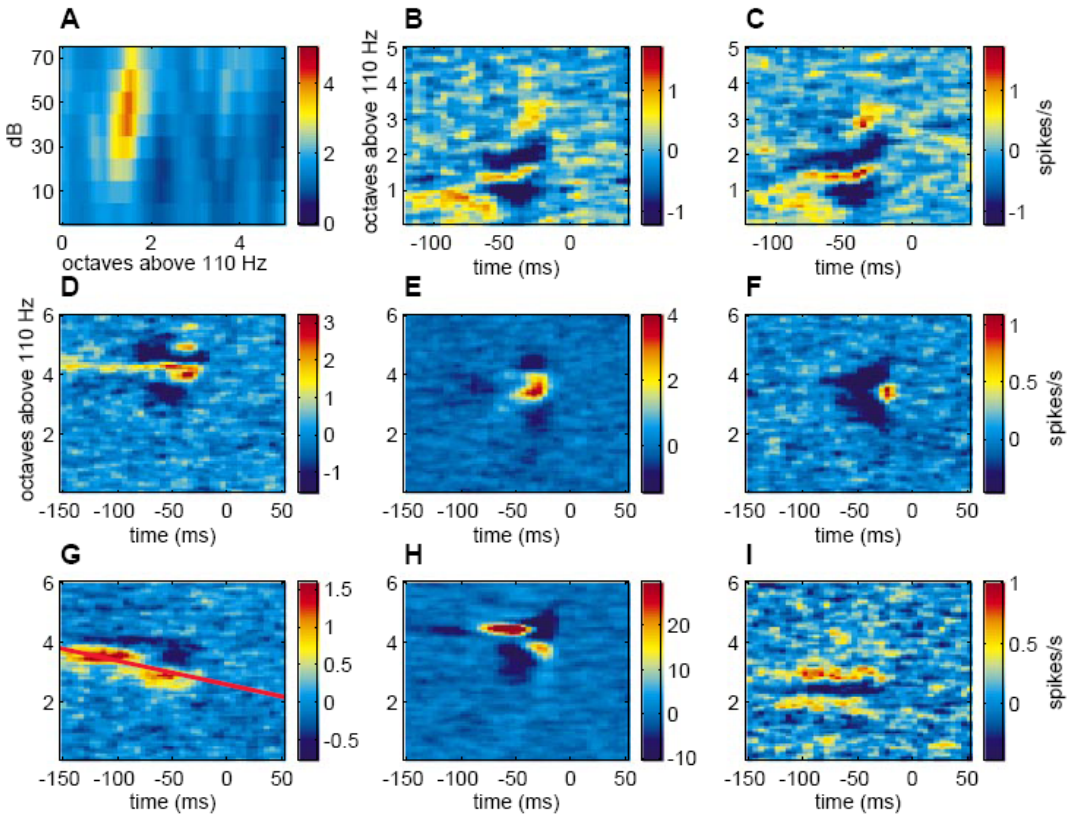


Figure 1-4 The spectrotemporal receptive fields of neurons in the primary auditory cortex of the awake primate show the patterns of sound features selected for by particular neurons. Note

that these auditory cortical neurons prefer transients in both amplitude (AM) and frequency (FM). (deCharms et al., 1998)

In sum, temporal modulation features are well and widely represented in auditory cortex, and acoustic stimuli containing dynamic features rather than stationary features seem to be a better trigger for auditory neurons. Large numbers of auditory neurons have a specific characteristic receptive field, preferring a certain combination of AM and FM features in the incoming auditory sounds. In this sense, they are acting as an ensemble of 'temporal modulation feature detectors', and sounds are decomposed and encoded in parallel in 'feature detector arrays'.

1.1.3 Speech recognition with temporal modulation features

As discussed previously, a complex sound is well characterized by a spectrogram, which contains all the detailed stationary and dynamic features that discriminate different sounds. This idea was also the basis of the speech recognition field back in the second World War, when the Sonograph was the main instrument in speech research by virtue of its detailed spectrographic portrait of the acoustic signal and its apparently objective and comprehensive description of all the details of the signal. However, a speech signal is not just an acoustic signal, in that not all the fine details in the spectrogram are required for humans to understand speech. In other words, we humans only extract certain relevant acoustic information from the speech signal for further comprehension, and therefore a detailed and precise description of speech is redundant to some extent. For example, we can tolerate many kinds of distortion in

speech, such as reverberation, noise influence, gap inserting, etc., and note that the spectrograms of this distorted speech are very different from the original spectrogram. This 'perceptual invariance' confirms the redundancy of the information contained in the spectrogram.

Then, what are the acoustic features most relevant to speech recognition? This question is the key aspect of human speech recognition studies. In late 1930s, Homer Dudley and his colleagues at Bell Labs developed the channel vocoder (1939), which passed the speech spectrogram into twenty or fewer channels and modeled the production of a speech signal as the filtering of a source signal by these filters. The resultant energy fluctuation patterns from these filters were extracted and transmitted for re-synthesis at the target unit. They found that high intelligibility could be obtained by keeping only the fluctuations below 20 Hz, indicating that linguistic information contained in a speech signal is actually encoded in relatively slow AM features (below 20 Hz), and that fine spectral details are not required. These results open the door to a new way of characterizing the speech signal and allow the construction of a framework focusing on the temporal evolution of coarse spectral patterns as the primary carrier of information within speech signals.

A modulation spectrum, which describes and quantizes spectral energy change over time, is computed by performing a spectral analysis of the signal's envelope, or the AM patterns of the signal, and provides a statistical characterization of the signal's temporal structure. For complex signals like speech signals, in which the energy change patterns or AM patterns differ for different frequency bands, the modulation spectrum needs to be calculated in band-limited frequency separately to

adequately reflect the signal's temporal changes (Greenberg & Arai, 2001; Greenberg, 2003). For example, as we described in the previous paragraph, the channel vocoder employed 20 or fewer frequency bands. Dudley et al. found that the information below 20 Hz in the modulation spectrums of these bands is critical for speech recognition. Since only a coarse representation of a spectrogram is needed for speech recognition, further studies investigated the minimum requirements to reach enough intelligibility by varying frequency bands and low-passing the modulation spectrum. They found that the number of bands can be decreased to 7, even to 4, depending on the type of speech tested, and low-frequency information in the modulation spectrum is most critical. FM cues were shown to be able to enhance speech recognition in much noisier listening conditions and more difficult tasks, compared to conditions where only AM cues were available (Zeng et al., 2005). In sum, speech recognition studies showed the important role of temporal modulation features in speech recognition and suggested that temporal modulation features need to be extracted and processed in the human brain for speech processing. Figure 1-5 shows the block diagram used in speech recognition studies to test the role of AM and FM cues in speech recognition and corresponding behavioral performances.

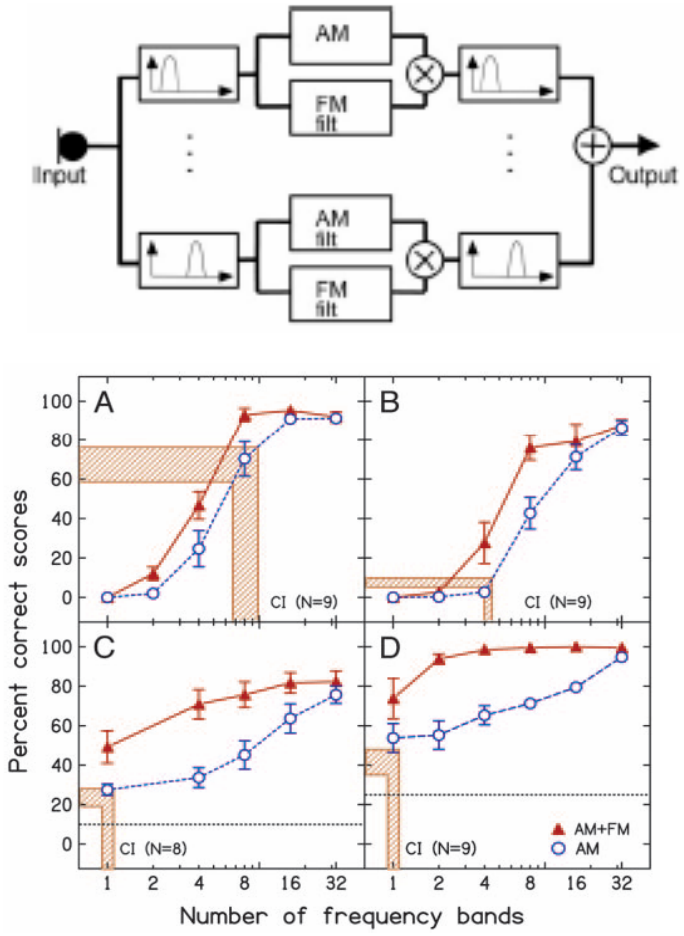


Figure 1-5 AM and FM cues contribute to speech recognition. Upper panel: signal processing block diagram. Lower panel: behavioral performances with AM and FM sounds; note the additional role when adding FM cues (Zeng et al., 2005).

1.2 Microscopic and macroscopic activities

Brain imaging techniques have brought exciting advances in neuroscience. Previous psychology methodologies were limited to the study of the input/output

relationship of the ‘black box’, which is actually an indirect estimation of innate activities and mechanisms. In contrast, these non-invasive brain imaging techniques enable us to open the ‘black box’ and observe the brain activity of normal humans directly, in real time. ,Ideally, by combining the advantages of hemodynamic and electromagnetic brain imaging techniques, we can observe high-quality spatiotemporal brain activities and possibly even achieve the goal of ‘reading the mind’.

On the other hand, data from non-invasive brain imaging seem to be a less direct reflection of brain activity than data from single-cell recordings. For example, typical measured electromagnetic signals require synchronous activation of 10,000-100,000 neurons (Wilson & Mcnaughton, 1987), and therefore brain imaging data are really at a more macroscopic level than traditional single-cell data.

Furthermore, even in the ideal case where we record human brain activity with both high temporal and high spatial resolution, could we really achieve the dream of ‘reading the mind’? Have we successfully gained comprehensive knowledge about animals in neurophysiological experiments by inserting electrodes in their brains and directly observing these neurons’ activities? What is the relationship between brain imaging data (macroscopic activities) and neurophysiological data (microscopic activities)? Are there any new perspectives and information that we could acquire from brain imaging in addition to its advantages in reflection of human brain activities? These questions are not insignificant issues and their answers will have deep influences on this field.

1.2.1 Overview of brain imaging techniques

In the last few years, the emergence of brain imaging techniques has helped neuroscience research enter a new era in which neuroscientists are able to see inside the living human brain as it performs various activities.. Brain imaging techniques include hemodynamic methods such as functional magnetic resonance imaging (fMRI) and positron emission tomography (PET), and electromagnetic techniques such as electroencephalography (EEG) and magnetoencephalography (MEG). Hemodynamic and electromagnetic brain imaging techniques have different spatial and temporal resolutions. Specifically, hemodynamic methods possess high spatial resolving power and therefore are suitable tools for determining anatomic organization and functional spatial distributions. Complementary to hemodynamic methods, electromagnetic methods are limited in their spatial resolution but virtually unlimited in their temporal resolution. Electromagnetic imaging techniques are temporally accurate to ~ 1 msec, a scale comparable to that of single-neuron spike patterns. Recently, numerous efforts have been made to integrate these two main types of brain imaging techniques and make use of their respective advantages, combining high-quality localization information provided by the hemodynamic methods with high-quality temporal data generated by the electromagnetic-based techniques in multiple ways (see the review by Horwitz & Poeppel, 2002).

Take fMRI as an example, conventional fMRI experiments explore brain activation during a particular perceptual or cognitive task, with the goal of determining which regions of the brain are involved in this task. Therefore, the underlying assumption is ‘functional specialization’, the expression of neuronal

activity in response to the specific perceptual features or cognitive processes under investigation, or motor behavior controlled by specialized cortical areas. Correspondingly, the imaging data analysis usually consists of 'cognitive subtraction', in which the activation associated with two or more mental states at all sampled brain locations is compared. The resulting 'activity difference map' is consulted, and the locations on the map with statistically significant larger activation values are regarded as answers. This assumption and the corresponding analysis, as the original motivations in the application of neuroimaging techniques in neuroscience, have yielded great knowledge and understanding in many facets of the field. However, as many researchers have realized, such an appealingly simple model has overlooked many possible brain mechanisms, for example, 'functional integration', which emphasizes the representation of information via interaction among different brain areas instead of being contained in specific areas. Other fMRI experimental designs and analysis methods have been reviewed by Friston (1997).

A main deficiency of the hemodynamic method lies in its coarse temporal resolution in measuring brain activities. As discussed previously, the small innate world, like the indefinite outside world full of changes, is dynamic all the time, and it is certainly necessary to include temporal information about brain activities. MEG/EEG are logical tools to employ for these purposes because of their excellent temporal resolution. MEG measures the magnetic field generated by neuronal current flow (Hamalainen, 1991, 1992). Because magnetic fields can pass undistorted through the skull, MEG is more spatially focal than EEG, which measures electric potentials mixed across many cortical areas. Due to the advantage of high temporal

resolution, when examining MEG/EEG activity, identifying the prominent activity (temporal peaks or troughs) in recorded signals is the main way to determine the neural correlates of certain cognitive tasks or mental states in time domain. Then source localization of these prominent responses can be computed by examining the corresponding magnetic field contour map at the critical time point or time window. A great deal of multi-disciplinary effort has gone into improving the 'dipole localization' algorithms for detailed source localizations, specifically for MEG, by employing various assumptions and incorporating data from other brain imaging techniques.

1.2.2 Neurophysiology: microscopic activities

The brain is an assembly of cells, each of which is a semiautonomous agent. Most neurons have similar structures and working mechanisms; they receive input at their dendrites and perform a wave-pulse conversion at their axons. A neuron acts on another neuron by sending a spike to its dendrite via the synapse, and the dendrite of the receiving neuron integrates the spike inputs it receives and transforms them to waves, which are transmitted to its axons, where the wave-pulse conversion is performed, and the 'all or none' spike with fixed height is generated. Therefore, the input and output of a single neuron are different in that the input is continuous dendritic waves whereas the output is discrete spike patterns and they have a nonlinear relationship.

Single cells are microscopic in terms of size as well as representation. Take a single sensory receptor cell as an example. This type of neuron, as the interface

between the world and the brain as well as the first stage of representation and processing, is truly microscopic in the sense that each neuron is 'seeing' only a small part of the stimulus, for example, colors in vision, frequencies in sound. Such fragmental representation forms a spatial pattern which is transmitted in parallel into the brain. When we further trace the information flow in the brain, we still observe such 'functional specialization' in neurons even at higher levels, and the only difference may be that they represent more abstract properties related to perception and cognition. This 'feature detector' role of neurons is a widely accepted concept and is also a reasonable organizing mechanism for efficient signal processing in the brain, similar to the division of work among machinery in modern factories. Single-cell recording studies have yielded enormously valuable information as to the functional structure of the brain, and most results support the framework of 'feature maps'. For example, what are neurons encoding? Which neurons are encoding the crucial features of the stimulus? What are their encoding schemes? However, the essential function of the brain is more complex than simply creating a stationary feature map, which begs the question: how are these features combined to form an integrated perception? This question is also known as the 'binding problem'. Take hearing as an example. The basic goal of hearing is to identify objects in the environment and to localize them in space (Yost, 1991). It is certainly a more complex task to identify and separate objects from background noise than to merely recognize the combination of all features of the sound, and there are many influences on such identification, an important one among which is context. Although in recent developments, more and more single cells are found to encode or be related to more

high-level and complex cognitive tasks, such as decision making, attention, context-dependent, preparation before vocalization, etc., the interpretations of these findings seem to ignore many secondary responses in these neurons and overemphasize the complex properties represented by these single cells.

Single-cell recordings provide us with a direct measurement of microscopic activities in the brain and lay a solid foundation for further understanding the mysterious, complex inner world. However, even when we fully understand each neuron, it is still not enough for us to directly relate them to the understanding of higher-level processing due to several concerns. First, they are redundant and related representations of a fragmented external world, and constructing objects from them is not at all a mere jig-saw puzzle that can be solved by simply seamlessly combining them. Secondly, those individual neurons are connected via excitatory or inhibitory synapses, and there are also interactions with neurons even in spatially remote areas. Such complex, dynamic local and global interactions make those fragmented microscopic representations too variable to directly infer from them what is happening at the level of the system. This is also a reason for the second response found frequently in most single neurons. Such failures of the reductionistic view exist in many fields (e.g., thermodynamics, sociology, geology, engineering, etc.), and therefore ‘system theory’ provides a more appropriate framework for describing macroscopic-level properties. The basis of this new perspective is that the structure of a system, that is, the relationships among its components, is often just as important in determining its behavior as are the individual components themselves. Importantly, the most critical function of brain, the integration of incoming information and the

construction of a single perception from them, may actually be reflected in its macroscopic activities, which is difficult to assess and infer from single-cell recording data.

1.2.3 Brain imaging: macroscopic activities

It has been suggested for a long time that a neuron population is a more reasonable candidate than the single neuron as processing blocks in the brain. This is a small step toward understanding the macroscopic activities of the brain, but it has raised a lot of technical difficulties. First, in order to record a population, many neurons need to be recorded simultaneously. Second, the recorded activity of a neuron population will be a large multi-dimensional data set, and a convincing encoding scheme which can be used to analyze such a data set has not been found. 'Temporal coherence' among neurons in the neuron group has received a great deal of support and may be a reasonable elementary representation mechanism at the macroscopic level. In addition, direct examination of data in corresponding multidimensional space is also a reasonable way to look at the output of neuron populations.

Brain imaging techniques have provided a natural means of recording brain activities at the macroscopic level. Since MEG is the main brain imaging method employed in my experiments and has reasonable spatial resolution and excellent resolution in time (~1 ms), I will mainly discuss MEG. The main source of the MEG signal is current flow in pyramidal cells' apical dendrites, and typical measured MEG

signals require the synchronous activation of 10,000-100,000 neurons (Wilson & Mcnaughton, 1987).

In keeping with traditional ERP studies, evoked responses averaged across many repetitive trials have received the majority of attention in MEG experiments. The underlying principles are that the brain will respond to the same stimulus condition with the same temporal response pattern, and in order to decrease the deleterious influence of background noise, averaging across trials could ideally recover the temporal response embedded across trials. Note that the assumption here is that the signal contained in each trial, specifically the temporal waveform, needs to be temporally phase locked to the trial onset, otherwise, the averaged response will be smeared due to temporal jitter. These averaged temporal waveforms can be regarded as the neural correlates of different stimulus or experiment conditions, and the prominent peaks and troughs occurring at certain points in the temporal waveforms will receive further attention in the form of detailed examination of properties such as amplitude, latency, and dipole localization. This method of analysis is straightforward and conceptually simple to understand because one of the main advantages of MEG over other brain imaging techniques is its temporal acuity, and it is this detailed temporal information that is the goal of MEG investigations. The most prominent evoked MEG response in the auditory field is the M100, an auditory response emerging 100ms after the onset of a sound stimulus that originates in the superior temporal gyrus (Lutkenhoner et al., 1998). The M100 can be found robustly in single subjects, and its two main parameters, amplitude and latency, have been found to be strongly correlated with spectral features of the acoustic signal. Further studies

suggest that, rather than being a pure auditory onset response, the M100 more closely reflects 'change detection' (Chait et al., 2005). Figure 1-6 illustrates an example of an M100 in a human subject.

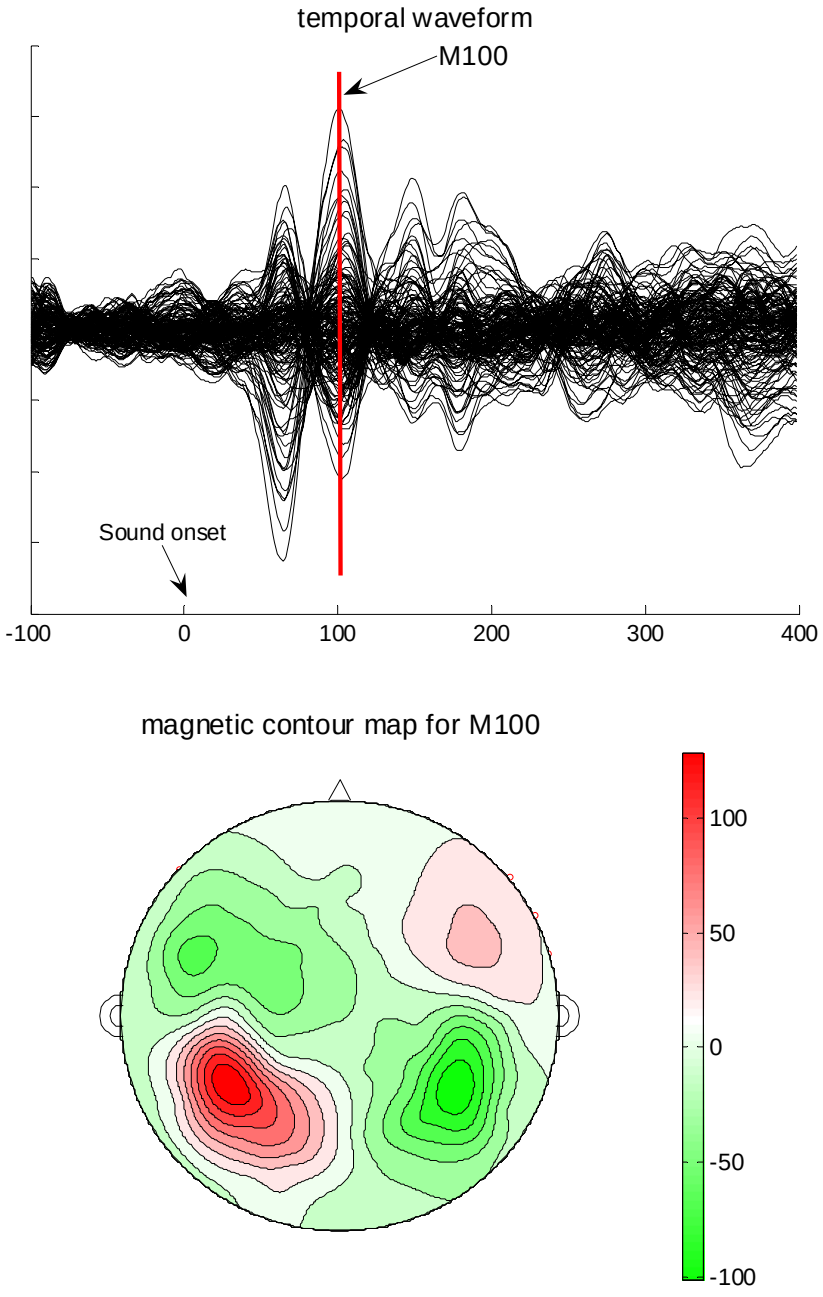


Figure 1-6 M100 temporal waveform and corresponding magnetic contour map, with red indicating the source and green indicating the sink.

The introduction of new signal processing methods and new perspectives on brain activity have brought important progress in the exploration of MEG responses. For example, spectrotemporal analysis, including induced wavelet analysis and evoked wavelet analysis, spectral analysis, temporal coherence analysis, principle component analysis (PCA), and independent component analysis (ICA), has been increasingly employed and has provided many important findings in this field. New algorithms have been developed by using more sophisticated signal processing theory and incorporating fMRI results have also enhanced dipole localization in MEG signals.

However, there is a much deeper question underlying all the possible analysis methods. What does the MEG signal really reflect? What are the main dimensions along which we should investigate elicited MEG responses? These are challenging questions without any obvious answers. We are put in a difficult situation: we are provided with an enormous data set, but how do we decode the information contained in it, and how do we understand the language that is being used there to transmit information?

A main feature of the MEG response is the dominance of continuous oscillations, especially at low frequencies (<20 Hz), and these oscillations can modulate and shape the corresponding evoked temporal waveform response to a large extent. For example, if we look again carefully at the averaged temporal response containing the M100 , we can see lots of other temporal peaks around 100ms. What do those peaks mean? Are they only minor noise fluctuations that can therefore be ignored? And if

we look at those temporal peaks channel by channel and explore their magnetic contour maps, we find that those additional peaks/troughs around the M100 are not random changes, but are closely related to the M100 and have the same origin. It is much easier to understand these characteristics of the waveform by regarding them as consecutive peaks/troughs of the same oscillating signal; the M100 is just the most prominent peak. In other words, traditional ERP analysis, by focusing on only one big peak at a certain time, misrepresents the complexity of the system, leading to misconceptions about how it actually works. The information contained in the MEG signal is not wholly conveyed in the form of large peaks at specific time points, but is transmitted continuously in dynamic oscillating waves that are slowly modulated by incoming stimuli or mental states. Figure 1-7 illustrates this oscillation in an evoked auditory MEG response.

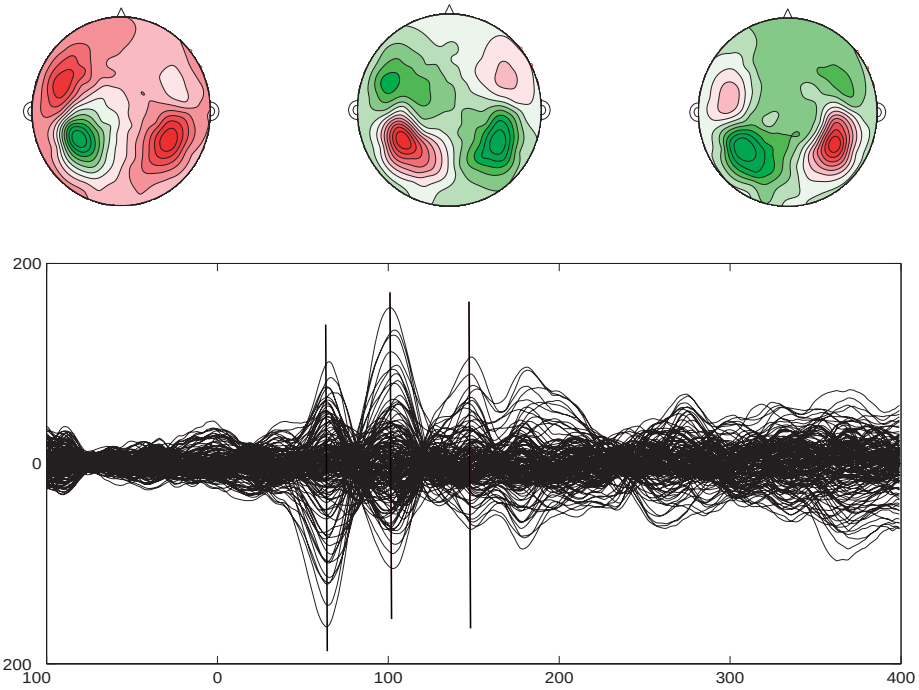


Figure 1-7 Oscillations in an evoked temporal MEG response to a 50msec 1-kHz pure tone presented at 0 msec. The three contour maps in the upper panel correspond to the three temporal peaks below, indicated by lines. Note other peaks around M100 and their similar origin.

The more reasonable way to look at an MEG signal is to regard it as waves oscillating at many frequencies and to examine changes in the properties of these oscillations. This view suggests that MEG responses are generated by the superposition of evoked oscillations with different frequencies (Makeig et al., 2002), and these different frequency bands have been proposed to play different functionally significant roles (Hari et al., 1997). In other words, this view differs dramatically from the view underlying traditional evoked temporal response analysis in that it considers oscillations at different frequency bands, rather than large, discrete peaks, to be the main information elements in MEG responses. This is a reasonable and logical way to look at macroscopic activities for several reasons. First, macroscopic activity is a reflection of system state activity, which consists of complex and dynamic patterns that are the result of numerous reciprocal interactions between excitatory and inhibitory neuron groups, and of information repeatedly flowing back and forth, both spatially and temporally. The main result of such a complex, dynamic system is oscillations. Secondly, oscillations provide a natural temporal coherence or means of temporal grouping because they reflect synchrony among underlying neuronal activity. And MEG/EEG signals actually reflect mass activation from the synchronization of large numbers of neurons. Thirdly, oscillations have properties

similar to those of our mental states—dynamic, continuous and integrative—and thus are reasonable candidates for representation of our innate world.

Neural oscillations have received wide interest recently, specifically in the field of systems neuroscience, and many studies have found chaotic and rhythmic activity patterns in the nervous system. Walter Freeman (1975, 2000) proposed that chaos is important for flexibility in nervous system responses, enabling the dynamic system represented in the cortex to change the attractor it approaches based on changes in incoming stimuli or internal mental states. Rodolfo Llinas (1988, 2000), based on years of research in the thalamocortical system, proposed that the gamma band (~40 Hz) plays a critical role in integrating content and context, and in embedding the external world in the internal world to construct one single entity—the self. In addition, there arises a new theory to explain the event-related potential (ERP) (note that the M100 is an example of an ERP), the oscillatory phase-resetting model, which suggests that oscillations in the theta (4~8 Hz) and alpha (8 ~ 13 Hz) frequency ranges undergo a significant phase resetting in response to the presentation of a stimulus (Makeig et al., 2002; Gruber et al., 2005), leading to the big peak observed in the ERP, for example, the M100. Figure 1-8 illustrates the phase alignment across trials that could account for the observed N1 and P1 components.

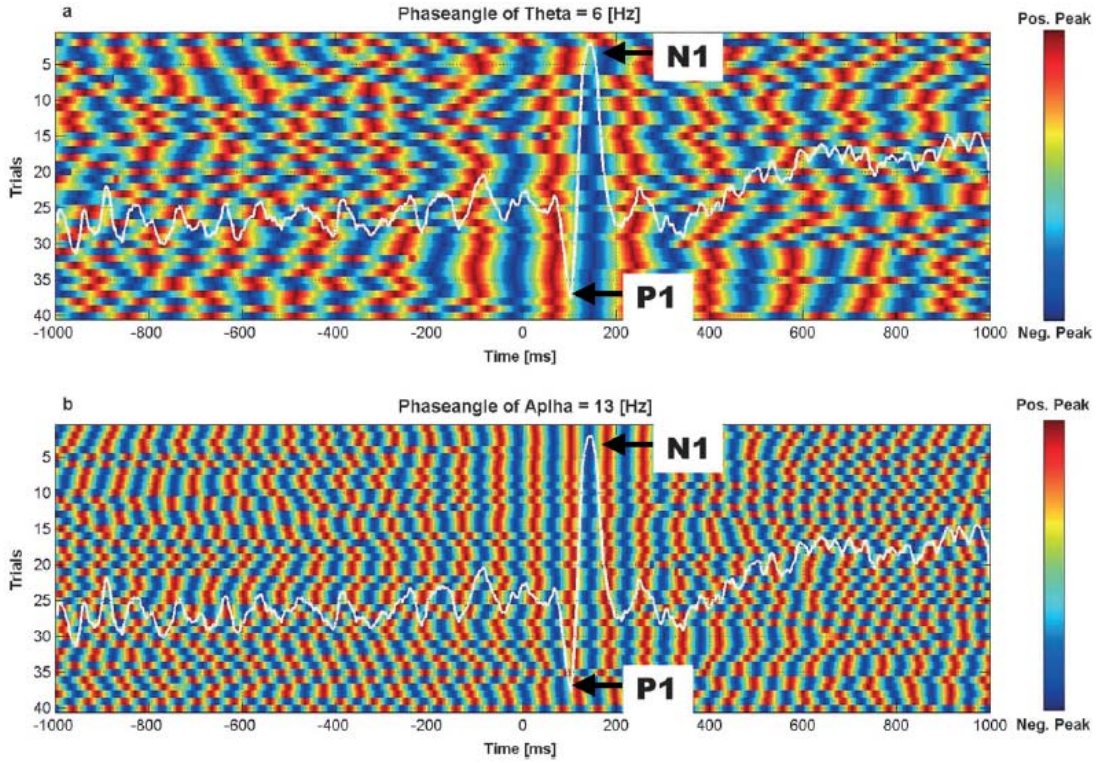


Figure 1-8 Phase alignment in theta band (upper panel) and alpha band (lower panel) across trials. Note that alpha phase synchronization after the stimulus onset shows better correspondence with the ERP components (while line) than theta phase alignment (Gruber et al., 2005).

1.2.4 Links between microscopic and macroscopic activities

In response to a dynamic and rich outside world, single cells perform their responsibilities by extracting specific features and representing this information via spike coding. Neurons early in the processing hierarchy hand in the information they have acquired to neurons at later stages, which fire spikes to encode more complex features. Single neurons are microscopic in the sense that they are only seeing a fragmented part of the external world, and they therefore cannot be directly related to our integrated mental state, which requires a mechanism of representation that is

unified, systematic, and of critical relevance to behavior. It is not difficult for us to believe that the neural correlates of our 'mental state' should be a systematic property emerging from the activity of enormous numbers of neurons that does not depend on any one or even several neurons, but is yet related to all of these microscopic cells. As we discussed previously, the activity recorded with MEG/EEG provides an efficient way for us to observe at least part of the activity of the system, and the prominent features of this macroscopic activity are oscillations at multiple frequency bands.

As discussed by Makeig (2002), studies of brain electrophysiology are dominated by two extreme subfields, single-cell spike histograms and ERP studies. These two methodologies are isolated from each other by differences in spatial scale, subjects recorded, and in part by modeling based on a simple averaging method. Meanwhile, the link between the timing of neuron spikes and the dynamics of the ERP remains largely mysterious. To achieve the goal of bridging this gap, investigators explore the intermediate stages between the two fields, specifically the dynamic information contained in single EEG/MEG trials, hoping to make progress toward understanding the dynamic consistencies between these fields in brain electrophysiology.

Numerous efforts have been made to try to understand the neural mechanisms of oscillations and to model the macroscopic activities reflected in fMRI and MEG/EEG experiments in terms of the activity of an underlying neural population (see review by Horwitz et al., 1999). Because fMRI results mainly convey spatial information, it is relatively straightforward to explain the macroscopic activities reflected in fMRI data, especially the spatial map, in terms of activation of functionally specialized neurons. Meanwhile, it is a more challenging and difficult task to create links between

microscopic activities and macroscopic activities in the temporal scale, due to the rich dynamics and dominant oscillations in MEG/EEG signals, and neither the mathematical structure of neural oscillations nor their functional significance are understood yet. The key questions for understanding the source or the origin of the observed patterns of neural oscillation are: Are single cells generating these patterns, or are they a property of a large neural system? Are these oscillations an innate property of the neural system, or are they the result of outside input?

At the single-cell level, collective oscillations in cortical neurons have been documented for several years and have been argued to be related to higher-order sensory representations (Basar, 1998); however, this idea continues to be hotly debated, because it is difficult to link oscillations to behavior. Therefore, when attempting to account for macroscopic oscillations in terms of the oscillatory activity of single neurons, a direct linear explanation seems to be a poor prospect.

Synchronizing input to neurons from sensory code is a straightforward solution to produce oscillations, and it is also the main solution to 'binding problems'. As early as 1974, Milner speculated that rhythmic activity patterns resulting from synchrony among neurons could play a role in binding parts of a perceptual pattern into a whole object from the environment. Gray and Singer (1989) later confirmed this speculation and proposed perceptual binding in terms of gamma-band synchronous temporal correlation among neurons. Temporal coherence is regarded as an additional coding dimension for building internal representations. Note that such an explanation more or less depends on the outside inputs.

Complex neuron communication along various spatial and temporal scales is another prominent explanation of oscillations; it does not directly depend on outside stimuli to account for the observed patterns, and it allows for the generation of innate oscillations as observed in brain background activity. Cortical structures have a wide range of intrinsic mechanisms that could generate synchronous activity. Inhibitory interneurons might be particularly important because of their ability to effectively entrain cortical neurons, thus making them good candidates for generating synchronized oscillations. In an effort to link ERP responses from MEG/EEG experiments to neural oscillations, it has been suggested that they are generated by the superposition of oscillations. Specifically, as noted previously, it is the phase resetting in low frequency bands (alpha, theta) in response to the presentation of a stimulus that leads to the large peaks seen in the waveform (e.g., P1, N1, M100). This perspective is very different from the more naïve view that regards these prominent peaks as the accumulation of activity in underlying neuron groups at a certain point in time. This view also reasonably explains the discrepancy between the long latency of these peaks in ERP and the short latency of onset spikes in single cells; the response resulting from the phase resetting of these long temporal windows (corresponding to low frequency bands) will occur relatively slowly. In a modeling study by David et al. (2005) that attempted to investigate the mechanism that shapes evoked MEG responses, the researchers constructed a neural mass model of hierarchically arranged areas using three kinds of inter-area connections (forward, backward and lateral), and studied how event-related dynamics depend on extrinsic connectivity. They found that adding backward connections could produce damped oscillations in the evoked

response, and adding bilateral connections could introduce damped oscillations as well as phase locking among areas.

In sum, we need to understand macroscopic activities from a somewhat different perspective, regarding them as system states resulting from the complex interactions of underlying microscopic activities. Correspondingly, the way we investigate those activities, specifically the MEG/EEG response, should be different from the stimulus-triggered spike pattern analysis used in single-cell data. Oscillation is a more elementary information block in macroscopic activities, and the aim is to study the dynamics in these oscillating signals in response to different stimulus conditions and mental states. The macroscopic information gained with MEG/EEG allows us to study the complex system from a systematic perspective, and the macroscopic-level activity is a more relevant neural correlate of behavior. Although macroscopic signals are concomitant with underlying neuronal activation, they cannot be simply and directly interpreted in terms of these microscopic activities.

1.3 Tracking sound dynamics in animals: neurophysiological studies

1.3.1 General approaches and neural encoding schemes

Temporal modulations are fundamental components of species-specific communication sounds, and therefore have been widely studied in neurophysiological experiments. The aim is to understand how these wide ranges of time-varying features in sounds are represented at the single-cell level. Neural representations of two main types of temporal information have been mainly investigated: temporal

precision to temporary transients such as stimulus onset, and the ability to follow sustained repetitive transients such as click trains. The first inquiry is related to the ability to detect abrupt changes in the environment, and the latter inquiry is related to the ability to perceive continuous dynamic structures in the outside world. Neurons exhibit paradoxical responses to these two types of temporal dynamics (Elhilali et al., 2004): on the one hand, many neurons have been demonstrated to have remarkable temporal precision of spikes in response to stimulus onset and other transients in single trials; on the other hand, these neurons fail to follow sustained repetitive stimuli beyond 20 Hz. This is called the resolution-integration paradox. Different stimuli have been used to investigate neural responses to both continuous and sudden changes in the environment. For example, tone onset and dynamic dots have been employed to study the transient response; click trains, amplitude- and frequency-modulated tones and noise have been widely used to study sustained tracking performance in neurons.

The investigation of sustained tracking performance is more relevant to understanding the mechanisms underlying speech processing and thus will be the emphasis in this thesis. The main analysis method used to quantify tracking performance is to calculate the 'vector strength' of the spike pattern in terms of the stimulus modulation rate. Note that the assumption behind this analysis is that the sustained temporal information in the stimulus is explicitly represented in single cells, and thus is the 'explicit temporal coding'. There are other possible neural coding schemes that can theoretically be employed by cortical neurons, for instance, rate

coding, which changes the spiking rate according to the varying modulation rate in the stimulus.

1.3.2 Tracking simple sound dynamics

The dynamic stimulus with the simplest modulated format is sinusoidally modulated tones or noise. The rationale for using this kind of stimuli is that, as shown by the Fourier theorem, a dynamic signal can be regarded as the sum of sinusoidal waves with different periods, corresponding to different frequencies. By studying neurons' representations of these 'atomic dynamic signals', we can gain fundamental knowledge about how fast and selective the modulation rate in the stimulus can be and still be tracked by the neurons. However, we should note that probably the performance of tracking complex stimuli probably cannot be fully inferred from the results for these atomic dynamic signals due to their complexity and nonlinearity.

The neural representation of temporal modulations begins at the auditory periphery, where the auditory nerve fires in a phase-locked fashion to pure tones of up to several kHz, and to the envelope of amplitude-modulated tones at modulation rates above 1 kHz. The precision of this temporal representation decreases at later stages along the ascending auditory pathway. A possible reason for this decay is that neurons at later stages receive converging inputs and perform 'temporal integration', where more integrative and complex properties are represented, which does not require precise preservation of all temporal transients.

In a series of studies by Xiaoqin Wang's group investigating neural representation of temporal modulations in unanesthetized marmoset auditory cortical neurons,

various types of temporally modulated stimuli were employed (Liang et al., 2002; Lu et al., 2001; Wang et al., 2003). In a study using narrow-band and wide-band click trains, they found two largely distinct populations of neurons: one with traditional stimulus-synchronized discharge, which is an 'explicit temporal encoding' scheme, and the other with non-stimulus-synchronized discharge, which employs 'implicit rate coding'. The 'temporal coding' neuron group could represent click stimuli with ISI longer than 20 ms, corresponding to a modulation rate of up to approximately 50 Hz, and the 'rate coding' neuron group could represent the same class of stimuli with ISI shorter than 20 ms corresponding to a modulation rate above 50 Hz. These two observed groups of neurons, encoding sequential stimuli in distinct ways, complement each other by representing a wide range of temporal dynamics (corresponding to a range of modulation frequencies) in the stimulus, thus providing neuroscientists with evidence that helps explain the capability to perceive temporally modulated sounds across the wide range of time scales demonstrated in behavioral experiments in both humans and animals. Wang's group also explored these two kinds of coding in auditory cortical neurons in response to sinusoidally amplitude- (sAM, with frequency fixed) and frequency- modulated (sFM, with amplitude fixed) tones. They found that the majority of neurons in A1 of awake marmosets showed similar selectivity for certain modulation rates in both sAM and sFM stimuli, indicating that the selectivity was a temporal rather than spectral phenomenon and confirming that temporal modulation is a general property represented in the cortex. Another important conclusion, arrived at by summarizing results across populations of A1 neurons, is that A1 is maximally synchronized to a temporal frequency of ~8

Hz, indicating that the low-frequency temporal modulation information, the crucial temporal scale for speech and melody recognition, is explicitly temporally encoded in spiking patterns, and the faster temporal transients, such as formant transients, may be implicitly coded by the neuron discharge rate.

Temporal modulation is a prominent feature in natural sounds, and interestingly, it has also been shown that adding co-modulation across different frequency bands can facilitate the detection of tones in noise by humans, a phenomenon known as co-modulation masking release (CMR). In a study by Nelken et al. (1999), they showed that co-modulation improved the ability of auditory cortical neurons to detect tones in noise, thus providing important evdience about the neural property underlying the behavioral CMR phenomenon. Another interesting observation from multi-electrode recordings in monkey auditory cortex is stimulus-induced gamma oscillations in local field potentials (Brosch et al., 2002). Although this result is not directly related to tracking sound dynamics, it provides some neurophysiological evidence for the macroscopic oscillations introduced in the previous section.

1.3.3 Tracking complex sound dynamics

Complex sounds contain more rich dynamic structure than simple sounds, and therefore the results from experiments using complex sounds, especially sounds with speech-like temporal modulation features, can contribute in a more direct way to our understanding of speech processing or natural sound processing in human brains. Additionally, because of the complexity of their temporal features and the nonlinearity of the auditory cortical system, the representations of complex sounds

are also more deeply embedded in some abstract format in the recorded signal, and such encoding schemes could not be estimated and inferred easily from the results for simple sounds.

'Ripple sounds', a class of acoustic stimuli composed of broadband frozen noise or harmonic tones with various spectrotemporally modulated envelopes, are one type of complex sound in terms of their rich temporal structures, and they have the advantage of being easy to manipulate. Another important property of this kind of sound is that it is made up of broadband signals and, like natural sounds, contains temporal modulations in both amplitude and frequency. Therefore ripples are more similar to natural sounds than are simple dynamic sounds, which contain temporal modulations in either amplitude or frequency. By reverse-correlating neurons' responses with the temporal modulations in the stimulus, spectrotemporal response fields (STRFs) can be computed, which indicate each neuron's preferred spectrotemporal features. In a study by Elhilali et al. (2004), they separately investigated the STRFs associated with envelope and fine structure dynamics in ferret A1 area, corresponding to slow and fast spectrotemporal modulations, respectively. Note that the product of these two features is a ripple sound. This scheme was used for two main reasons: First, the researchers wanted to investigate the 'resolution-integration' paradox by employing stimuli containing simultaneous slow and fast temporal modulations and to explore the relationship between these two aspects of cortical responses. Secondly, as mentioned previously, in speech recognition studies, it has been found that slow temporal modulation, which corresponds to the envelope dynamics here, is crucial to speech recognition, while fine structure dynamics are

helpful in much noisier environments. Therefore to explore their simultaneous representations in auditory cortex, and especially their relationship, would be very useful. Interestingly, by using these kinds of stimuli, they solved the paradox. They found that most neurons can track the envelope and fine structure modulations simultaneously and manifest a dual response, although this is contingent on the cell being driven by both of these modulations. This is a very interesting result, and it suggests that the failure to observe neurons firing phase-locked to fast temporal modulations is probably the result of using stimuli with only one type of temporal modulation . In other words, by employing more complex dynamic sounds that also have slow modulation features, neurons could express their ability to track the fast temporal modulations which could not be observed in simple dynamic sound experiments. Sounds with more natural dynamic structures seem to be a more efficient trigger of neurons' ability to represent temporal modulation features, and these findings underline the importance of using dynamic sounds with more natural temporal structures.

The responses of single cortical neurons to natural communication sounds have also been widely studied, which is a further step toward understanding the representation of speech in the brain. The auditory cortex plays an important role in the perception of complex sounds, especially species-specific communications. It has been found that many neurons in the primary auditory cortex of marmoset monkeys respond more robustly to conspecific vocalizations compared to the same vocalizations played in reverse, although physically the stimuli contained similar temporal structures.

A reasonable research assumption underlying the investigation of natural communication sounds is that neurons in the auditory cortex may shape their 'characteristic response area' in terms of the statistical structure of natural sounds. Based on this view, many studies have analyzed temporal structure across wide ranges of natural sounds and have designed artificial acoustic stimuli containing the same temporal modulation features found in natural sounds. They investigated the cortical responses to these speech-like stimuli, and in order to determine whether they had succeeded in creating efficient triggering stimuli, they tested whether the features they had included were in fact the main features that auditory neurons were analyzing and processing (Woolley et al., 2005).

Another more direct way of exploring the encoding of natural communication sounds in cortical neurons is to study whether spike trains in a single neuron can discriminate different communication sounds. In other words, can single neurons fire spiking patterns that encode the temporal structure of natural sounds? Based on the temporal information contained in spike trains, can we decode the natural sound the animals have heard? This is really an intriguing inquiry and has been tested in different types of animals. A study by Machens et al. (2003) exploring auditory receptors' responses to individual songs in grasshoppers showed that short segments of single spike trains from one auditory receptor suffice to rapidly discriminate the songs of conspecific grasshoppers, provided that a time resolution of a few milliseconds is maintained. Strikingly, the researchers found distinct temporal scales for the success of this observed discrimination ability. Specifically, the spike trains need to be explored using a temporal window of about 5ms to get good discrimination

ability; smaller or larger temporal windows will decrease discriminability. In addition, the system needs at least 200ms–400 ms to begin to demonstrate discrimination ability. The first short time scale corresponds to the sustained discrimination temporal scale to be used to encode continuous dynamics in songs, and the second long time scale corresponds to the time that the receptor neurons require for integration of information before reliable tracking and discrimination arise. A similar scheme has been used in natural sound discrimination in songbirds; there again, two distinct temporal scales for such discrimination were found, although the specific values in each range are different (Narayan et al., 2006).

In sum, we can reach several conclusions based on neurophysiological studies of tracking complex sounds. First, natural sounds, although containing more complex dynamic features than simple sounds, making them more difficult to control,, may nonetheless be the more efficient and relevant stimuli ensemble for activating and exploring the auditory system. The reductionist perspective to infer the complex from the simple may not be applicable here. Secondly, at a minimum, we should investigate the auditory system using sounds that contain the main temporal structures and properties of natural sounds. For instance, we should design sounds that have simultaneous amplitude and frequency modulation instead of one or the other, as has been the in most studies, considering most natural sounds contain concurrent amplitude- and frequency-modulation features. Thirdly, the most relevant representation and encoding mechanism for complex natural sounds lies in the temporal information of responses, and therefore using temporal scales to quantify responses is an important dimension of analysis dimension that we should consider.

1.4 Tracking sound dynamics in humans: brain imaging studies

1.4.1 fMRI/PET studies

The cortex has been proposed to be the primary site for temporal information processing. A fundamental issue concerns the property of temporal information. Is time a general property that is represented in all areas in a relatively explicit way, or is it a specific property that is processed in a specialized area? Hemodynamic imaging techniques, including fMRI and PET, could provide important information about functional spatial localization and have been extensively used in various types of experimental designs to find a temporal information processing center.

Another important issue is the lateralization of function found in auditory cortex. Many efforts have been made to try to link this lateralization to temporal information processing. Current studies are trying to understand the nature of such lateralization, for example, language processing in left hemisphere and music processing 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. More specifically, temporal features are predominantly processed in the left hemisphere and spectral features in the right hemisphere (Zatorre & Belin, 2002).

A related but different approach to explaining functional lateralization is the 'asymmetrical sampling in time' hypothesis (Poeppel, 2003). In this theory, both hemispheres work together to analyze auditory signals on multiple temporal scales,

with the relevant scales being short (~25–50 ms) and long (~200–300 ms) windows. These two scales have been shown to be the most important and behaviorally relevant time scales in information processing across many sensory domains. The lateralization emerges from having different principal temporal scales in the two hemispheres; information processed on longer timescales is routed predominantly to high-order right hemisphere cortices, whereas information processed on a shorter timescale primarily projects to the left cortices. This theory explains the observed functional lateralization in terms of this difference in the basic auditory information processing mechanism. On this view, the dominance of fast temporal transients (e.g., formant transitions, etc.) in language is what leads to the left-lateralized activation; the left hemisphere works at a faster temporal scale. Conversely, the dominance of relatively slow temporal changes in music results in these types of stimuli being routed primarily to the right hemisphere, which works at this longer time scale.

This hypothesis has been tested and confirmed to some extent in an fMRI experiment (Boemio et al., 2005) using a stimulus ensemble with a parametrically varying segmental structure affecting primarily temporal properties. Specifically, the 9-s auditory stimuli were designed by concatenating short-duration narrowband noise segments. The bandwidth and center frequency of each segment was chosen to match speech formants. Two types of segments with different local spectrotemporal variations were designed: in one type, the frequency remained constant throughout the signal; in the other, frequency was swept linearly upward or downward at random. Therefore, the ensembles of stimuli differ in two dimensions and have two controlled acoustic properties: the temporal structure introduced by varying segment length and

the local spectrotemporal properties introduced by using different types of segment structure. The stimuli were effective at activating auditory cortex selectively and robustly. Boemio et al. found that both early and higher-order auditory cortical areas are exquisitely bilaterally sensitive to temporal structures, and local spectrotemporal structures are differentially processed within the superior temporal gyrus. In addition, in higher-order superior temporal sulcus, slowly modulated signals preferentially drive the right hemisphere. These findings support hemisphere lateralization in STS, possibly due to neuron populations working at different time scales (short or long) in STG differentially targeting STS, with the right hemisphere receiving afferents carrying information processed on the long time scales, and the left hemisphere receiving information processed on short time scales.

In addition to mere spatial localization information, many researchers also look at the sound-evoked, blood oxygen level-dependent signal response with fMRI, which could provide additional temporal information. In an fMRI experiment by Seifritz et al. (2002), they used independent component analysis (ICA) in a hierarchical combination of spatial ICA and temporal ICA to blindly decompose the evoked blood oxygen level-dependent signal into its constituent spatiotemporal sources. They found that the temporal auditory response could be decomposed into a transient and a sustained component, which predominated in different portions—core and belt—of the auditory cortex. These findings suggest that in analyzing incoming sound, the higher-order auditory areas (the belt area here), which show sustained activity, process and analyze auditory information continuously, accounting for the detailed

analysis of sound information. The low-order areas (the core area here) may play a role in the detection of a new sound object via transient responses.

In another fMRI experiment, Harms and Melcher (2002) investigated the influence of noise burst repetition rates on properties of fMRI signals. They found a systematic change in the form of fMRI response rate-dependencies from low to high levels in the auditory pathway. Specifically, at lower stages (the inferior colliculus and the medial geniculate body), response amplitudes increase with increasing stimulus rate, and the response shape remains relatively unchanged. In auditory cortex, the response wave shape changes dramatically with increases in stimulus rate—low rates elicit a sustained response, whereas high rates elicit an unusual phasic response. 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. In other words, the investigators proposed that the neural correlates of auditory perception lie in corresponding population activities in auditory cortex that are reflected in the shape of the fMRI response. They designed new fMRI studies to try to understand exactly which properties of sound determine the cortical response shape by employing various kinds of temporally modulated stimuli. They found that the temporal envelope is the characteristic of sound that determines whether the fMRI response is phasic or not. They further proposed that the fMRI response wave shape reflects the segmentation of the auditory environment into meaningful events and could be a window allowing us to observe underlying neural activities. These results reconfirm that the macroscopic activities reflected in brain imaging responses play significant roles in perception and provide novel information

about population activity, adding meaningfully to our understanding of information processing in the brain.

Recently, a new research field has emerged to try to directly assess how well a mental state can be reconstructed from non-invasive measurements of brain activities in humans by employing new statistical pattern recognition analysis (see review by Haynes & Rees, 2006). The fundamental difference between the assumptions here and those of traditional fMRI analysis is that here it is assumed that the information contained in the recorded macroscopic activities is not spatially localized, but is embedded in the whole spatial map. Correspondingly, the way to seek the relevant activity pattern is not to determine which specific brain region is statistically significantly involved in a certain task or temporal course, but to regard the whole spatiotemporal pattern as providing information. Such a pattern-based multivariate analysis approach could in principle allow considerable increase in the amount of information that can be gained compared to traditional strictly location-based univariate approaches. The pattern-based approach allows the investigation of other information-carrying mechanisms, for instance, 'across-place' joint representation.

These techniques and assumptions have contributed a great deal to the question of 'mind reading'. Although most successful examples are in the visual domain, they could be extended to audition. For example, recent work demonstrates that pattern-based decoding of BOLD contrast fMRI signals can successfully predict the perception of low-level perceptual features (e.g., orientation, direction of motion, etc.), whereas conventional approaches cannot (Kamitani & Tong, 2005; Haynes & Rees, 2005a). In an fMRI experiment investigating binocular rivalry, researchers

found that the perceptual fluctuations could be dynamically decoded from fMRI signals in the visual cortex by training a pattern classifier to distinguish between the distributed fMRI response patterns associated with the dominance of each monocular percept (Haynes & Rees, 2005b). Although these findings were not in the auditory domain, they at least in principle convey several important points about the essence of macroscopic activities. This is especially true considering that there are many common processing mechanisms between different sensory domains: the neural correlates of much higher-order perception or mental state are in a complex and dynamic format, and they are spatially distributed. We need to develop broader methods of analysis in order to understand system-level activities. Also, representational mechanisms in multiple dimensions should be given more emphasis.

1.4.2 MEG/EEG studies

Electromagnetic brain imaging methods, including MEG and EEG, have been used widely in recent investigations of brain mechanisms due to their high temporal resolution. The popularity of these techniques is not surprising in view of the fact that many experimental findings suggest that temporal information is crucial for our understanding of the brain's working machinery. MEG/EEG techniques are especially attractive and appropriate tools for investigating questions in the auditory domain because of the dynamic nature of this field.

As with research in neurophysiology, at the beginning, studies examining the representation of auditory temporal features in the human brain use simplified versions of sounds that have atomical temporal modulation features, for instance,

click trains, amplitude- and frequency-modulated tones, amplitude-modulated noise, etc. Motivated by results in single-cell recordings that suggest that temporal information is explicitly and temporally represented in spike trains, researchers are also expecting to find some temporal representation of temporal modulations in elicited EEG/MEG responses. The simplest example is to study EEG/MEG responses to click trains, and if activity in the brain can track the click train, the recorded EEG/MEG signals should be phase-locked to the consecutive clicks, as has been found in neuron spike trains. It is true that we find an elicited response in the temporal waveform for each click when the click trains are slow enough (SOA > ~ 500 ms), however, we do not find such explicit tracking when click trains become fast. The reasons for this are as discussed before: the recorded macroscopic signals were not simply the sum of underlying microscopic activities; they have their own characteristics, oscillations, and examining only the elicited temporal response may not be an appropriate way to decode such representations. In most MEG and EEG experiments in humans using AM or FM stimuli or click trains, the auditory steady state response (aSSR), an elicited response that has the same frequency as the corresponding stimulus modulation frequency, is the main approach to examine AM and FM representations in these recorded EEG/MEG signals. The principle underlying aSSR is that if responses show tracking of temporally modulated sound, in the corresponding spectrum we should observe the frequency component at this modulation frequency. A second principle of aSSR is that analysis in the spectral domain is better than analysis in the temporal domain because the target signal may be deeply masked by background activities across many frequencies and may

therefore be difficult to decode from such a mixed signal. aSSRs have been found for stimulus modulation rates up to 200 Hz (Ross et al., 2000, 2005; Picton et al., 2003). The principles of aSSR analysis are illustrated in Figure 1-8.

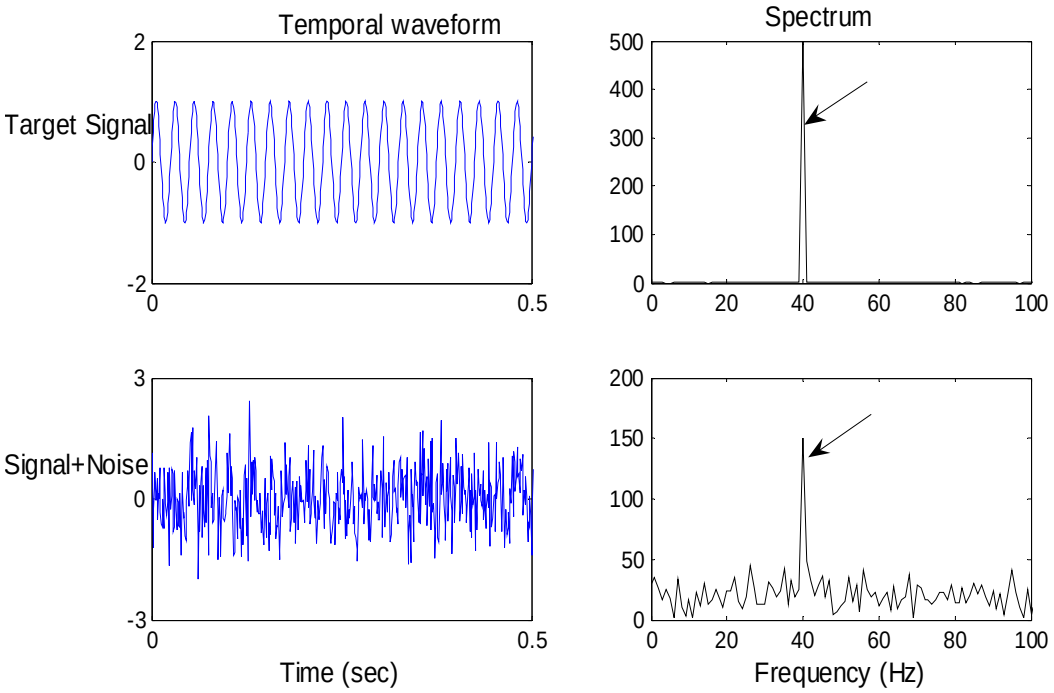


Figure 1-9 Principles of aSSR analysis. Analysis in the spectral domain is more robust to background noise than is direct examination of the temporal waveform by detecting the spectral peaks at corresponding target modulation frequency (black arrow).

Interestingly, aSSRs have been found to be maximal when the stimulus modulation frequency is around 40 Hz; these findings hold for various stimulus types and across sensory domains. From the perspective of systems neuroscience, it has been suggested that the elicitation of the 40-Hz brain rhythm by transients in the sensory input could account for the maximum at 40 Hz in aSSR, because when the

stimulus modulation rate is 40 Hz, the elicited gamma-band response for each transient will be in phase with and enhance each other, and the cumulative result isa strong aSSR at 40 Hz (Galambos et al., 1981). Such stimulus-elicited gamma oscillations have also been directly observed in the spectrograms of EEG signals, adding further support to this explanation (see Boemio thesis, 2003). In addition, a comprehensive study of the aSSR using pure AM sound (Ross et al., 2000) systematically examined the effects of stimulus properties (modulation frequency, carrier frequency) on the aSSR (amplitude and phase), suggesting that properties of the 40-Hz brain oscillation are modulated by the properties of incoming sensory stimuli.

A more important aspect of this explanation of aSSR is that the observed EEG/MEG signals were actually superpositions of oscillations with various frequencies, and any representations will exist in terms of the changing properties of these oscillations, such as amplitude and phase. These waves modulate their own properties in response to external stimuli. For example, here, the sustained acoustic transients (amplitude or frequency) elicit specific gamma-band oscillations, leading to the observed aSSR at the stimulus modulation frequency in the spectral domain. As for the M100 example introduced previously, it is the phase resetting of low-frequency bands (theta and alpha) after detecting a sound stimulus that contributes to the emergence of this prominent auditory response.

Finding neural correlates of natural communication sounds in EEG/MEG signals is more challenging due to several unsolved problems. Although in principle, we should find different activities for different sound stimuli, practically, we lacked

sufficient knowledge about the dimensions that we could depend on to robustly decode and discriminate these complex sounds. There are several MEG experiments that could provide inspiration.

An MEG study by Ahissar et al. (2001) explored whether they could find neural correlates of temporal information in speech sentences, motivated by speech recognition studies showing that temporal information is most crucial for speech intelligibility. In a speech comprehension task, they tested subjects using temporally compressed speech at varying compression ratios, corresponding to different levels of intelligibility, and recorded responses from auditory cortex using MEG. By performing principle component analysis (PCA), they found that the first three principle components could account for more than 90% of the variance within the sensor array. They tested whether these principle components, containing time-varying information, could account for the different degrees of intelligibility of the compressed speech stimuli. They found that the first principle components have a time course corresponding to the temporal envelopes of the stimuli when analyzed in the spectral domain, where the similarity between the prominent frequencies and phase locking could account for the intelligibility of the corresponding compressed speech. Specifically, stronger similarity indicates high intelligibility and vice versa. In other words, they found a striking link between temporal response in brain activities and behavioral performance. Their results support the notion that behaviorally correlated representations in macroscopic activities are complex and need to be examined in a more abstract way across time and space, and they emphasize the key role of temporal information in such representations.

Another very interesting MEG study, by Patel and Balaban (2000), explored whether music-like tone sequences could be reliably tracked and represented in human cortical activity. Motivated by previous aSSR studies, they used amplitude modulation of unfamiliar, long tone sequences to try to label stimulus-related MEG temporal responses. They successfully demonstrated that the temporal patterns of recorded MEG responses tracked the pitch contour of tone sequences, with the accuracy of tracking increasing as tone sequence became more predictable in statistical structure. Specifically, it is the phase of the elicited aSSR at the amplitude-modulation frequency that reliably tracked the tone sequence. These results also support the significance of temporal information in the MEG signal for representing complex dynamic stimuli. Furthermore, as shown in neurophysiological studies of ferrets using ripple sounds, it seems to be most efficient for the brain to track the features of complex dynamic sounds that are variable in both envelope and fine structure simultaneously. Another important indication from this study is that when examining the macroscopic-level activity reflected in MEG and EEG signals via spectrotemoral analysis, phase information should receive equal attention with frequency power as a candidate for information-carrying elements of the signal..

1.5 Summary

From a reductionist perspective, a complex system can be understood in terms of relatively simple components. Such a linear view is appropriate and useful at the start of research in especially complex systems, and this view has yielded a great deal of information about how the brain works. Two main directions in brain research follow

from reductionist views: understanding complex information processing in terms of simple information processing, and understanding system activities in terms of microscopic activities.

As mentioned previously, temporal modulations are dominant features in both the external and internal worlds, and therefore understanding their representation in the brain is a crucial undertaking. Large numbers of studies have employed stimuli with the simplest versions of temporal modulation features to explore this question and try to infer the mechanisms underlying complex and natural temporal feature processing. However, due to the complexity of the auditory system, there may be a bottleneck for such approaches.

The reductionist perspective on analysis of brain imaging activities may overlook many other formats in which information is embedded in the macroscopic-level output, leading to misinterpretation or overemphasis of what are perhaps epiphenomenal response components. For example, peaks and troughs in the temporal waveforms of MEG signals have traditionally been regarded as the main information carriers relevent to certain stimulus conditions or cognitive tasks, whereas new evidence has shown that they are actually an epiphenomenal component resulting from other representational mechanisms. Overinterpreting these components underestimates the complexity of the system and will mislead research in this field.

Efforts should be made to overcome the confusion resulting from reductionist views, in terms of both understanding natural auditory information processing and employing new perspectives to examine macroscopic activities. This thesis covers several inquiries into these unsolved problems in the auditory domain using MEG.

The main question is, how are the temporal structures of a complex sound represented and processed in human auditory cortex?

Although many different types of stimuli could be used to investigate these questions, complex sounds with speech-like temporal features and natural human speech sounds will be used. Several new methods of MEG analysis based on various perspectives of MEG activities will be introduced in order to try to decode the information hidden in the dynamic and complex macroscopic signal. Two important beliefs guide the whole framework of this thesis: first, brains are constructed to process natural, unified stimuli in the best and most efficient manner; secondly, we need to find an appropriate way to decipher the information buried in the observed MEG responses.

Chapter 2: Tracking simultaneous acoustic AM and FM features

2.1 Introduction

A fundamental issue in auditory neuroscience concerns the nature of the computation that transforms the raw sensory signal into a representation that is useful for auditory tasks (Smith & Lewicki, 2006). Complex sounds, especially natural sounds, can be parametrically characterized by many acoustic and perceptual features, one among which is temporal modulation. Temporal modulations describe changes of a sound in amplitude (amplitude modulation, AM) or in frequency (frequency modulation, FM).

Amplitude modulation (AM) and frequency modulation (FM) are two important physical aspects of communication sounds, corresponding to the independent envelope and carrier dynamics of a sound. They are found in a wide range of species-specific vocalizations for both animals and humans (Doupe & Kuhl, 1999). In speech recognition studies, acoustic envelope (i.e. AM) cues were shown to be crucial to speech intelligibility (Drullman et al., 1994, Shannon et al., 1995). Analogously, Zeng et al (2005) have shown that acoustic carrier (e.g. FM) cues significantly enhance speech recognition performance even under noisy listening conditions, in contrast to AM cues, which enhance recognition only under ideal listening conditions.

Furthermore, these temporal modulation features are known to be encoded in the auditory system. Numerous neurophysiological studies in animals have indicated that precise timing information is preserved throughout the ascending auditory pathways (Heil, 1997; Oertel, 1997, 1999; Eggermont, 2002; Philips et al., 2002; Elhilali et al., 2004; Rose & Metherate, 2005). Using reverse correlation techniques, it can be shown that the response properties of auditory cortical neurons are dominated by transient changes in both amplitude and frequency, reflecting their selectivity for AM and FM features in the stimulus sounds (deCharms et al., 1998; Deprieux et al., 2001; Miller et al., 2002; Elhilali et al., 2004). Interestingly, the reverse approach, which makes the theoretical assumption that that the auditory system's encoding mechanisms are shaped to represent natural sounds in the most optimal and efficient way, predicts a preponderance of AM and FM response patterns in the receptive fields of auditory cortical neurons (Lewicki, 2002; Klein et al., 2003).

Physiological responses to both AM and FM sounds have been widely studied in non-human species (Schreiner & Urbas, 1986, 1988; Eggermont, 1994; Gaese et al., 1995; Heil & Irvine, 1998; Liang et al., 2002), as well as in humans, using electroencephalography (EEG) and magnetoencephalography (MEG) (Rees et al., 1986; Ross et al., 2000; Picton et al., 2003), fMRI (Giraud et al., 2000), and intra-cranial recordings (Liegeois-Chauvel et al., 2004). There is also a rich psychophysical literature of behavioral responses to modulations (Zwicker, 1952; Viemeister, 1979; Moore & Sak, 1996). However, it is still debated whether AM and FM sounds are processed using the same or different mechanisms and pathways (Saberi & Hafter, 1995; Moore & Sek, 1996; Patel & Balaban, 2000, 2004; Dimitrijevic et al., 2001;

Liang et al., 2002). Animal studies show that cortical neurons can fire phase-locked to amplitude modulated sounds up to tens of Hertz (Schreiner & Urbas, 1986, 1988; Eggermont, 1994; Gaese et al., 1995). However, rate coding instead of temporal coding has been observed for higher rates (Lu et al., 2001). In addition, there is a high degree of similarity between cortical responses to AM and FM stimuli (Liang et al., 2002), suggesting at least some shared representation of temporal modulations by cortical neurons (Wang et al., 2003).

Neuroimaging techniques have also been extensively used as a tool to study the processing and representation of temporal modulation features in human auditory cortex. Functional Magnetic Resonance Imaging (fMRI) and intracortical recording experiments have revealed sustained cortical responses to AM and FM sound stimuli that vary in magnitude and shape as the stimulus modulation rates increase above 10 Hz (Giraud et al., 2000; Harms & Melcher, 2002; Hart et al., 2003; Liegeois-Chauvel et al., 2004; Haller et al., 2005). In most MEG and EEG experiments on humans using AM or FM stimuli, the auditory steady state response (aSSR), an elicited response with the same frequency of the corresponding stimulus modulation frequency, is the main approach to examining AM and FM representations. aSSRs have been found for stimulus modulation rates up to 200 Hz (Ross et al., 2000, 2005; Picton et al., 2003), consistent with the stimulus-synchronized discharge (or the temporal coding) observed in animal studies. Cumulatively, these results reveal that cortex apparently encodes incoming auditory signals by decomposing them into envelope and carrier (Smith et al., 2002).

Although AM and FM have been widely studied in both animals and humans, the auditory systems are often probed with either AM or FM stimuli. Natural sounds, however, contain simultaneously modulated envelope and carrier frequencies (both AM and FM). Therefore, instead of manipulating the envelope or carrier dynamics separately, the auditory cortex may be probed using stimuli with both dynamic envelope and carrier.). In other words, AM and FM always co-occur and are inseparable acoustic features of an auditory object, and therefore the auditory system should be able to co-track them to achieve 'perceptual unity' of the incoming sound. Note that 'co-tracking' refers to a combinational encoding of AM and FM features, and it differs from simultaneous tracking, in which the resultant neuronal activity is simply a sum of the two separate tracking signals for AM and FM, respectively. The latter example is an EEG experiment (Dimitrijevic et al., 2001), which employed independent amplitude and frequency modulation (IAFC) stimuli with relatively higher modulation frequencies (above 80 Hz) and found independent aSSR responses for both AM and FM. There have been at least two examples of such 'co-tracking' found in auditory systems. Elhilali et al. (2004) have shown that single units from primary auditory cortex (AI) in ferrets lock to both slow AM and FM modulations and to the fast fine structure of the carrier (up to carrier frequencies of a few hundred Hz). Patel & Balaban (2000, 2004), using MEG, investigated the processing of sinusoidally amplitude modulated tone *sequences* (co-modulation of both envelope and carrier where the slow frequency modulation is periodic but not sinusoidal), and showed that the phase of the aSSR at the envelope modulation frequency tracks the tone sequences, i.e. the carrier changes. These results indicate a relation between the

representation of AM and FM features processing in human auditory cortex and a possible co-tracking mechanism.

How might auditory cortex co-represent envelope and carrier dynamics simultaneously? Modulation encoding is one important possibility. Modulation is a way to describe stimulus dynamics, such as the AM and FM signals, it is also a very important method to embed a general information-bearing signal into a second signal, or to co-represent two signals. AM, FM, and related modulation schemes are widely used encoding techniques in both nature and electrical engineering. One class of modulation encoding is AM, in which the modulation signal is used to modulate the amplitude of another signal, called the carrier. Another important class is phase modulation (PM), in which the signal needing to be transmitted modulates the phase of the carrier signal. FM is a generalized PM, in which the signal needing to be transmitted modulates the time derivative of the carrier phase (which is also equal to the carrier's instantaneous frequency). There are also a wide variety of other modulation encoding methods used for other radio transmission applications, including single sideband, single sideband with suppressed carrier, and double sideband with suppressed carrier. These encoding schemes have the advantage of efficiently transmitting signals even in the presence of noise. These encoding schemes can be used to transmit signals even in the presence of noise, whether electromagnetically in the radio band, or neurally in the auditory system (Oppenheim & Willsky, 1997).

In the Fourier domain, modulated signals have distinctive signatures, which may be easier to detect and decode than their time-domain versions. A narrowband carrier

appears as a single peak in the spectrum at $f_{carrier}$, the carrier frequency. The modulations due to either pure AM or pure PM appear as sideband frequency patterns in the spectrum. Specifically, the spectrum will have an upper sideband at $f_{carrier} + f_{modulation}$ and a lower sideband at $f_{carrier} - f_{modulation}$ (often accompanied by additional, lower power, sidebands at more distant frequencies), and different modulation-type signals (e.g., AM, FM, and PM signals) can be distinguished based on the phase relationships among those sidebands and the carrier. At least one example of modulation encoding is seen in human auditory cortex: at extremely slow frequency modulations (~0.1 Hz), the phase of the envelope modulation frequency aSSR tracks the carrier change, i.e. a form of PM encoding (Patel & Balaban, 2000, 2004). Whether other methods are used, and what method is used at higher frequencies, is largely unknown.

Figure 2-1a illustrates these basic concepts from the engineering encoding point of view. Figure 2-1b shows the hypothesized spiking activity corresponding to neural modulation encoding (third row: PM encoding; fourth row: AM encoding) of the considered stimulus with sinusoidally modulated carrier frequency (first row, FM) and amplitude (second row, AM). An ensemble of PM encoding neurons (third row) will produce an evoked neural PM signal similar to that shown in the middle of the lower panel of Figure 2-1a (obtained mathematically by low-pass filtering the spike train). Similarly, an ensemble of AM encoding neurons (fourth row) will produce an evoked neural AM signal similar to that shown in the middle of the upper panel of Figure 2-1a.

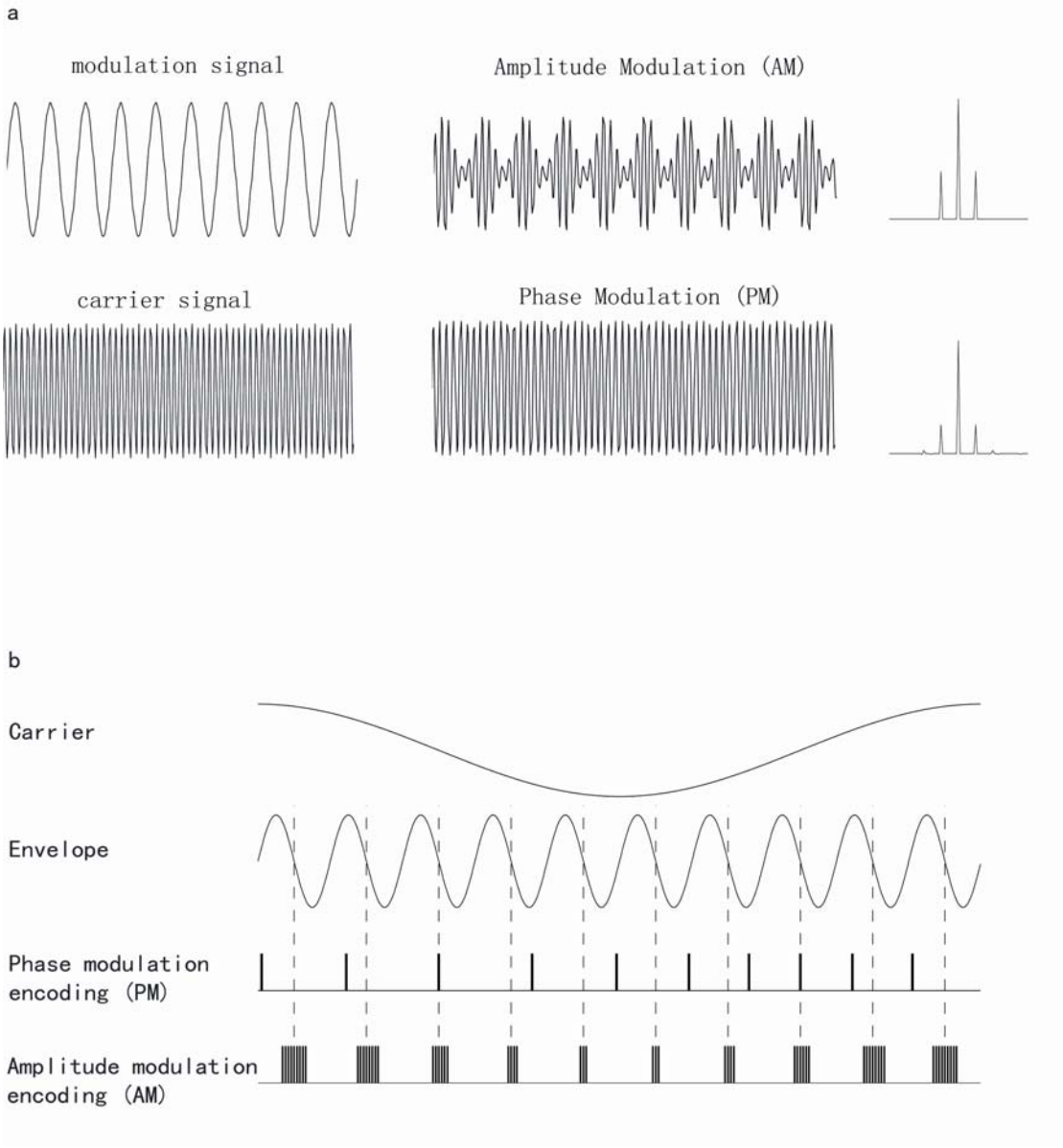


Figure 2-1 Modulation as an encoding method in engineering, and proposed neural mechanisms. a) A modulation signal modulates either the amplitude or the phase of the carrier signal to produce either an AM signal or a PM signal. Both signals produce a two-sideband pattern in spectrum (right). b) Possible neural modulation encoding mechanisms for AM encoding and PM encoding to simultaneously represent both stimulus carrier (first row, changes in stimulus carrier frequency) and stimulus envelope (second row, changes in stimulus amplitude) dynamics. A neuron employing PM encoding (third row) fires one spike per stimulus envelope cycle, as indicated by the dotted line, and the firing *phase* in each cycle depends on the instantaneous

stimulus carrier frequency. A neuron employing AM encoding (the last row) changes firing rate according to the instantaneous stimulus carrier frequency, while keep the firing phase within each cycle fixed (aligned with the dotted line).

The ability of auditory cortex to track stimulus dynamics via the aSSR is limited. The aSSR to AM sounds can be recorded with MEG from humans at stimulus rates up to ~100 Hz, with a large peak around 40 Hz (Ross et al., 2000); EEG responses follow to higher rates (see, e.g. Picton et al., 2003) but responses at those higher rates are not generated by auditory cortex. The aSSR at the modulation frequency, however, is generated only by neural temporal coding, whereas many neurons employ rate coding for rapidly modulated stimuli (Lu et al., 2001). Therefore, it is still not fully understood how - and how fast - auditory cortex can track a stimulus, particularly for stimuli modulated in both envelope and carrier, as is typical of most ecologically relevant signals.

We designed sound stimuli that were sinusoidally modulated in both amplitude (AM, at rate f_{AM}) and frequency (FM, at rate f_{FM}). These stimuli are a simplification of natural sounds containing simultaneous AM and FM, but their dynamics can be simply described by two frequency parameters: f_{AM} and f_{FM}. In turn, we can examine their representations in the human brain by checking the spectrums of the measured MEG responses at those frequencies related to these stimulus dynamics parameters (f_{AM}, f_{FM}). In addition, by varying these dynamics parameters, we can investigate coding transitions as a function of stimulus rate dynamics.

The present study was designed to address three questions: First, how does human auditory cortex represent or co-represent simultaneous AM and FM. Second, how fast can human auditory cortex track the carrier dynamics (FM). Third, is there any coding transition as the rate of carrier dynamics increases? To address these issues, we take advantage of the high temporal resolution of MEG, which has shown to be a method with outstanding sensitivity to record from human auditory cortex.

2.2 Materials and Methods

2.2.1 Subjects

Slow-FM Experiment

12 subjects (8 males) with normal hearing and no neurological disorders provided informed consent before participating in this experiment. The subjects' mean age was 25 and all were right handed. A digitized head shape was obtained for each subject for use in equivalent-current dipole source estimation.

Fast-FM Experiment

11 subjects (including several subjects from Slow-FM Experiment) with normal hearing and no neurological disorders provided informed consent before participating in this experiment.

2.2.2 Stimuli

Slow-FM Experiment

Nine stimuli were created, using custom-written MATLAB programs (The MathWorks, Natick, MA), with a sampling frequency of 44.1 kHz. The stimuli were sinusoidally frequency modulated tones with modulation frequencies **(f_{FM}) of 0.3, 0.5, 0.8, 1.0, 1.7, 2.1, 3.0, 5.0, 8.0 Hz** and frequency deviation between 220 Hz to 880 Hz. In addition, the entire stimulus amplitude was modulated sinusoidally at a fixed rate of **37 Hz (f_{AM})** with modulation depth of 0.8. All stimuli were 10 s in duration and shaped by rising and falling 100 ms cosine squared ramps. Each stimulus was presented 10 times. Figure 2-2 shows the spectrogram (higher panel), the spectrum (middle panel) and the temporal waveform (lower panel) of example stimuli, confirming that the stimulus sounds contain both sinusoidally modulated temporal envelope at f_{AM} (37 Hz) and sinusoidally modulated carrier frequency at f_{FM} (0.8 Hz and 2.1 Hz as examples drawn here). Because the frequency range of the carrier ranges from 220 Hz to 880 Hz, the stimuli have the broadband spectra shown in middle panel.

To ensure that subjects attend to the long stimulus sequences, 36 distracter stimuli were created and inserted into the experiment for subjects to detect. Those distracters were the same as the normal stimuli except single short-duration FM sweeps were inserted at random time in the stimulus. Subjects were instructed to press a button when detecting the distracter stimuli. Normal stimuli ($90 = 9 \times 10$) and distracter stimuli (36) were mixed and played in a pseudo-random order at a comfortable loudness level to subjects. Subjects performed the required task fairly well (average miss rate: ~ 3/36; average false alarm rate: ~ 1/36). The entire experiment was

divided into 4 blocks with breaks between them. Only the data for normal stimuli were further analyzed.

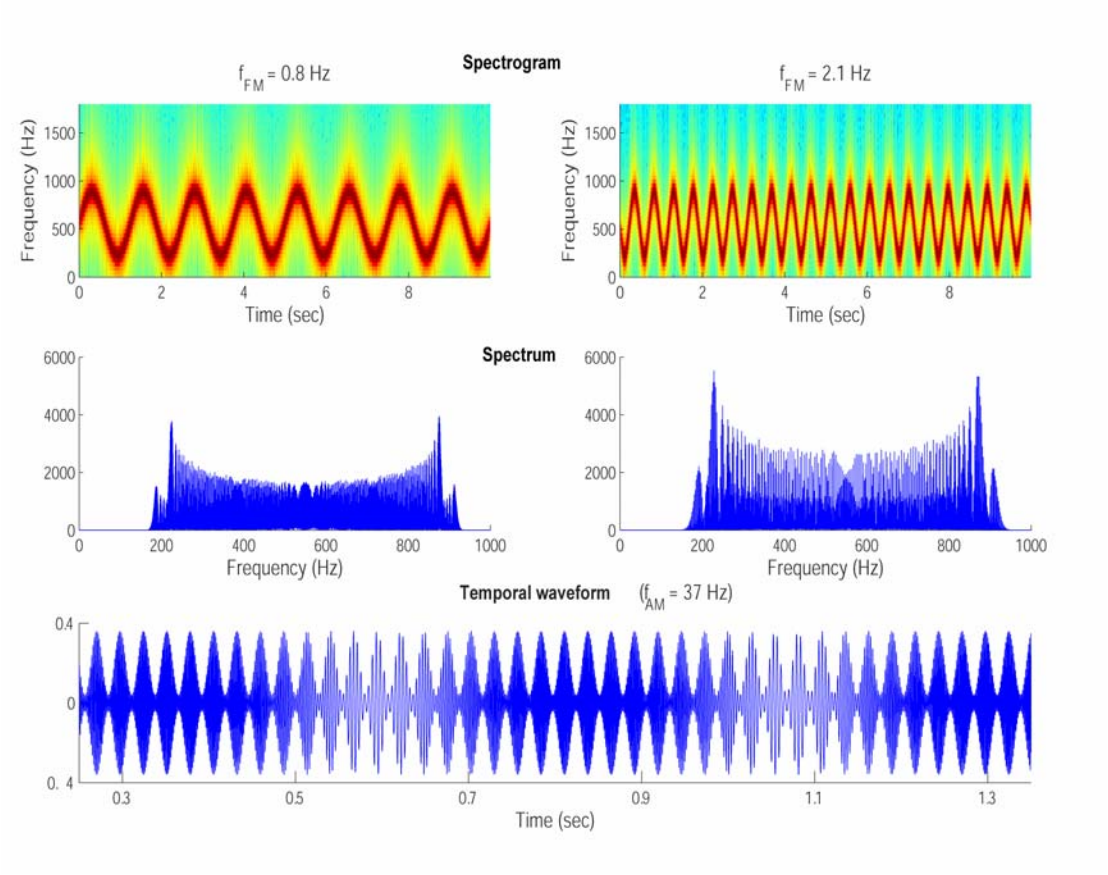


Figure 2-2 Slow-FM Experiment stimulus examples. Top: the spectrograms of stimuli with f_{FM} equal to 0.8 Hz and 2.1 Hz. The carrier frequency was modulated at a particular frequency (left, 0.8 Hz and right, 2.1 Hz), sinusoidally from 220 Hz to 880 Hz. Middle: the corresponding spectra of the stimulus examples in upper panel (left, 0.8 Hz and right, 2.1 Hz). Note that the spectra are broadband. Bottom, temporal waveform of stimulus with f_{FM} equal to 2.1 Hz. The envelope of the stimulus is modulated sinusoidally at 37 Hz. Only one segment from 0.2 sec to 1.4 sec is shown to let the 37 Hz AM be seen more clearly. The carrier change can also be seen here. The stimuli have both dynamic envelope (lower panel) and carrier (upper panel).

Fast-FM Experiment

The same custom-written MATLAB programs (The MathWorks, Natick, MA) was used to design another 9 stimuli with a sampling frequency of 44.1 kHz. The stimuli were also sinusoidally frequency modulated tones but with relatively higher modulation frequencies **(f_{FM}) of 2.1, 3.1, 5.1, 8.0, 10.3, 15.1, 20.1, 24.3, and 30 Hz** and frequency deviation between 220 Hz and 880 Hz. Note that the stimulus f_{FM} values overlapped between Slow-FM Experiment and Fast-FM Experiment in 4 values (2.1 Hz, 3.1 Hz, 5.1 Hz and 8.0 Hz) to check the robustness of results.

In addition, the entire stimulus amplitude was also modulated sinusoidally at a fixed rate of **37 Hz (f_{AM})** with modulation depth of 0.8. They were also 10 s in duration. Similarly, 36 distracter stimuli were created and inserted into the experiment for subjects to detect to ensure that subjects attend to the long stimulus sequences. These distracters were identical to the high-f_{FM} normal stimuli here except that single short-duration FM sweeps were inserted at random times in the stimulus. Subjects also performed the required detection task fairly well (average miss rate: ~ 4/36; average false alarm rate: ~1/36). Only the data for normal stimuli were further analyzed. Figure 2-3 shows the temporal waveform (higher panel), the spectrogram (middle panel), and the spectrum (lower panel) of two example stimuli, confirming that the stimulus sounds contain both a sinusoidally modulated temporal envelope at f_{AM} (37 Hz) and a sinusoidally modulated carrier frequency at f_{FM}. The stimuli also have the long-term broadband spectra shown in the lower panel.

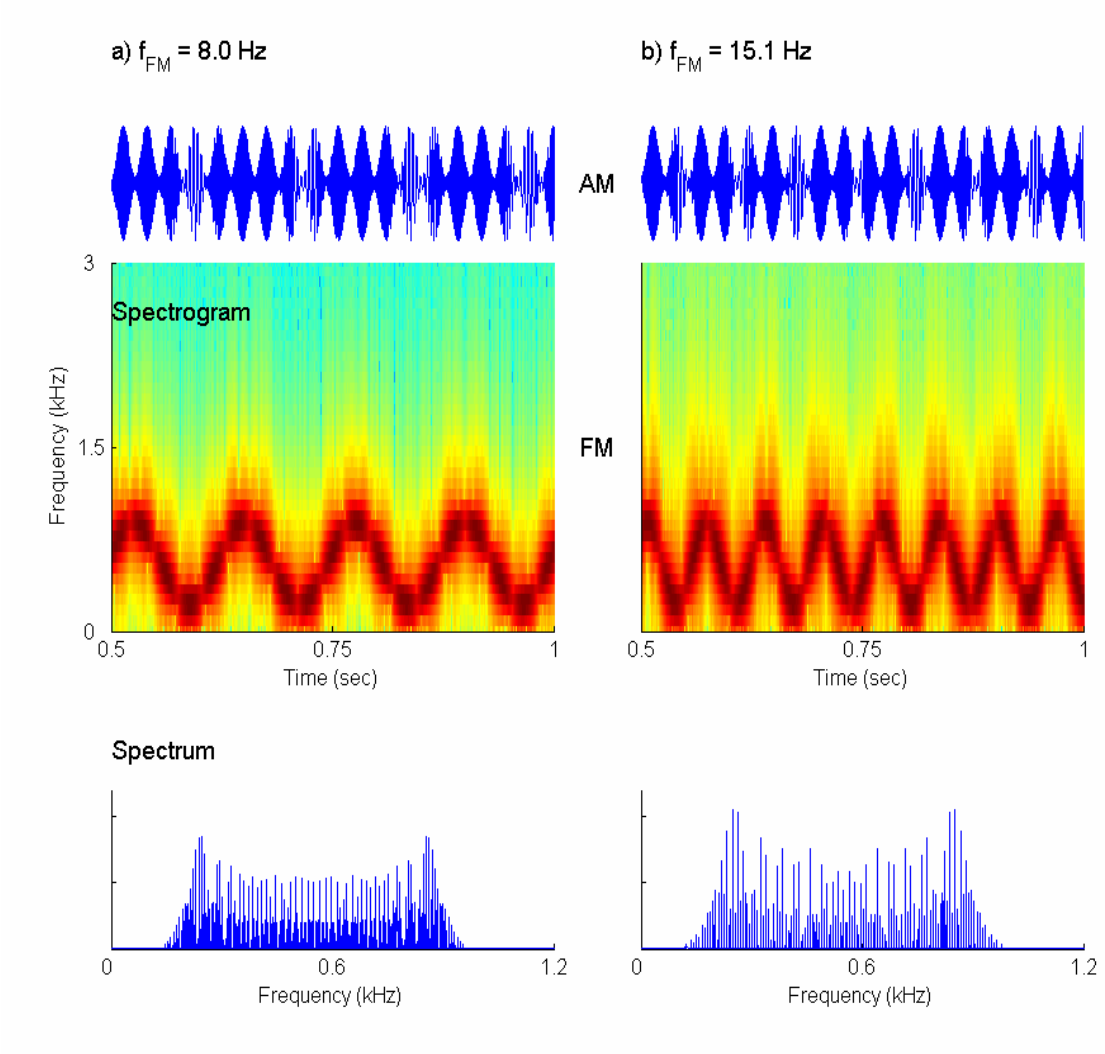


Figure 2-3 Fast-FM Experiment Stimulus examples with *f*FM of 8.0 Hz (a) and 15.1 Hz (b) respectively. Top: temporal waveform of stimulus. The temporal envelope was sinusoidally modulated at 37 Hz (*f*AM). Only one segment from 0.5 sec to 1.0 sec is shown to let the modulation be seen more clearly. Middle: the spectrogram of the same temporal segment (0.5 sec–1.0 sec) of the stimulus. Note the carrier frequency is also sinusoidally modulated (a, 8.0 Hz; b, 15.1 Hz) in the range from 220 Hz to 880 Hz. Bottom: spectrum of the stimulus (10sec duration). Note that the spectra are broadband.

2.2.3 MEG recordings

All experiment procedures were approved by the Institutional Review Board (IRB) of the University of Maryland. Neuromagnetic signals were recorded continuously with a 157 channel whole-head MEG system (5 cm baseline axial gradiometer SQUID-based sensors, KIT, Kanazawa, Japan) in a magnetically shielded room, using a sampling rate of 1000 Hz and an online 100 Hz analog low-pass filter, with no high-pass filtering. Each subject's head position was determined via five coils attached to anatomical landmarks (nasion, left and right pre-auricular points, two forehead points) at the beginning and the end of recording to ensure that head movement was minimal. Head shape was digitized using a three-dimensional digitizer (Polhemus, Inc.).

2.2.4 Data analysis

Data analysis has been done in Slow-FM Experiment and Fast-FM Experiment separately using the same data analysis procedures. First, auditory steady state (aSSR) responses were obtained by calculating the discrete Fourier Transform (DFT) of the concatenated responses from 10 trials (100 s = 10 $\times$ 10 s) for each of the 9 stimulus conditions (Slow-FM Experiment: f_{FM} = 0.3-8.0Hz; Fast-FM Experiment: f_{FM} = 2.1-30.0Hz), giving frequency resolution 0.01 Hz. These calculations were computed for all 157 MEG channels, all 9 stimulus conditions, and all subjects (Slow-FM Experiment: 12 subjects; Fast-FM Experiment: 11 subjects). In addition, the phase coefficients were adjusted with respect to the 60Hz hardware notch filter properties in

order to remove the phase shift introduced by the notch filter. These Fourier coefficients were stored for further analysis for each subject.

Phasor representation and Channel selections

For each channel, the steady state response (aSSR) at 37 Hz (f_{AM}) is parameterized by the DFT component's magnitude and phase at 37 Hz (f_{AM}). The result is a map of complex aSSR, i.e. a map of complex magnetic field values. An example of such a map can be seen in Figure 2-4b, where the complex magnetic field at each channel is represented by a phasor, i.e. an arrow with length proportional to the complex field magnitude and with direction given by the complex field phase (Simon and Wang, 2005). Then, the 50 channels per subject with the largest magnitudes across all the channels in both hemispheres at the 37 Hz (f_{AM}) modulation frequency were regarded as channels representative of auditory cortical activity and selected for further analysis, motivated by the positive relationship between tracking performance and response strength at f_{AM} found in an MEG experiment exploring representation of tone sequence in human auditory cortex (Patel and Balaban, 2004). The remaining channels were not further analyzed.

aSSR and M100 Equivalent-Current Dipole localization

To localize the neural source of the aSSR, the complex aSSRs corresponding to f_{FM} = 0.3 Hz in Slow-FM Experiment were analyzed to determine the best (least mean square) fit for a pair of equivalent-current dipoles (Simon and Wang, 2005). The resulting complex dipoles' positions, one in each hemisphere, are the estimates of

the source locations. These aSSR source locations are compared to the M100 source locations, estimated by the purely real version of the same algorithm. The M100 was measured in a pretest experiment, in which subjects were instructed to count the number of 1 kHz pure tones they heard. The M100 component is believed to originate in the superior temporal cortex on the upper bank of the superior temporal gyrus slightly posterior to Heschl's gyrus on the planum temporale (Lutkenhoner and Steinstrater, 1998). This direct comparison permits an analysis of the aSSR location without requiring magnetic resonance image (MRI). Such dipole localization was only performed in Slow-FM Experiment.

Sideband frequencies

Target sideband frequencies were defined for different f_{FM} as upper sideband ($f_{AM} + f_{FM}$) and lower sideband ($f_{AM} - f_{FM}$), leading to 18 (9 ×2) frequencies in both Slow-FM Experiment (upper: 37.3, 37.5, 37.8, 38, 38.7, 39.1, 40, 42, 45 Hz; lower: 36.7, 36.5, 36.2, 36, 35.3, 34.9, 34, 32, 29 Hz) and Fast-FM Experiment (upper: 39.1, 40.1, 42.1, 45, 47.3, 52.1, 57.1, 61.3, 67 Hz; lower: 34.9, 33.9, 31.9, 29, 26.7, 21.9, 16.9, 12.7, 7 Hz). The DFT amplitude and phase at every target sideband frequency were extracted for all 50 channels (selected specifically per subject), and for every stimulus condition, giving a 18 × 9 × 50 × 12 data set in Slow-FM Experiment and a 18 × 9 × 50 × 11 data set in Fast-FM Experiment (frequency × stimulus_condition × channel × subject).

Sideband amplitude matrix (A_{upper}, A_{lower})

We examined the presence of sideband patterns ($f_{AM} \pm f_{FM}$) in the spectra of the MEG signal, a distinctive signature of modulation encoding, by checking whether each specific stimulus condition (characterized by stimulus f_{FM}) induced significant spectral peaks at corresponding sideband frequencies ($f_{AM} \pm f_{FM}$) and not at other sideband frequencies. This analysis was performed for both Slow-FM Experiment and Fast-FM Experiment separately.

The amplitudes of a specific sideband frequency were examined for all 9 stimulus conditions and for all 50 selected channels, and the results were summed across 50 channels, giving a 9-value vector, which was then normalized by dividing by its mean. This 9-value vector represented the normalized elicited spectral power at this specific sideband frequency under all 9 stimulus conditions, so ideally, the maximum value will occur for the entry corresponding to the appropriate stimulus condition. The same procedure was followed for all sideband frequencies (9 upper and 9 lower sideband frequencies separately), giving two 9×9 matrices, corresponding to the upper sideband amplitude matrix (A_{upper}) and the lower sideband amplitude matrix (A_{lower}). In each amplitude matrix, the 9 rows represent the 9 different target sideband frequencies (in A_{upper}: $f_{AM} + f_{FM}$; in A_{lower}: $f_{AM} - f_{FM}$), and the 9 columns represent the 9 different stimulus conditions. Each element in the matrix represents the normalized spectral power at this specific sideband frequency (corresponding row) for a specific stimulus condition (corresponding column).

Encoding-type parameter α

Sidebands naturally occur for all types of modulation coding (including AM and PM). To help determine which modulation coding created the sidebands, an encoding-type parameter (α, defined below, ranging between 0 and 2π) was calculated to distinguish AM encoding from PM encoding. Both encoding mechanisms (see Figure 2-1b) elicit two sidebands, but with different phase relationships across the sidebands and carrier, characterized by the encoding-type parameter α (itself a generalized phase taking on values between 0 and 2π). The encoding-type parameter α is defined as

$$\alpha = (\theta_{upper} - \theta_{f_{\mathrm{AM}}}) + (\theta_{lower} - \theta_{f_{\mathrm{AM}}})$$

, where θ is the phase at that frequency, $upper = f_{\mathrm{AM}} + f_{\mathrm{FM}}$, and $lower = f_{\mathrm{AM}} - f_{\mathrm{FM}}$. AM encoding produces α near 0 (or 2π) and PM encoding produces α near π.

The mathematical derivation follows. For neural response carrier frequency f_c (identified with f_{AM}), neural response modulation frequency f_m (identified with f_{FM}), and modulation index *m*, this is shown for the neural response case of AM:

$$\begin{aligned} S_{\mathrm{AM}}(t) &= (1 + m\cos(2\pi f_m t + \phi_1))\cos(2\pi f_c t + \phi_2) \\ &= \cos(2\pi f_c t + \phi_2) + \frac{m}{2}\cos(2\pi(f_c + f_m)t + \phi_1 + \phi_2) + \frac{m}{2}\cos(2\pi(f_c - f_m)t + \phi_2 - \phi_1) \\ &= \cos(2\pi f_c t + \phi_2) + \frac{m}{2}\cos(2\pi f_{\mathrm{upper}} t + \theta_{\mathrm{upper}}) + \frac{m}{2}\cos(2\pi f_{\mathrm{lower}} t + \theta_{\mathrm{lower}}) \end{aligned}$$

.

Where we have set $\theta_{\mathrm{upper}} = \phi_1 + \phi_2$ and $\theta_{\mathrm{lower}} = \phi_2 - \phi_1$. Thus,

$\alpha_{AM} := (\phi_{\text{upper}} - \phi_2) - (\phi_2 - \phi_{\text{lower}}) = ((\phi_1 + \phi_2) - \phi_2) - (\phi_2 - (\phi_2 - \phi_1)) = 0$, which is also equivalent to $\alpha_{AM} = 2\pi$.

Correspondingly in the neural PM case,

$$\begin{aligned} S_{\text{PM}}(t) &= \cos(2\pi f_c t + \phi_3 + m\cos(2\pi f_m t + \phi_4)) \\ &= \cos(2\pi f_c t + \phi_3)\cos(m\cos(2\pi f_m t + \phi_4)) - \sin(2\pi f_c t + \phi_3)\sin(m\cos(2\pi f_m t + \phi_4)) \\ &\approx \cos(2\pi f_c t + \phi_3) - m\sin(2\pi f_c t + \phi_3)\cos(2\pi f_m t + \phi_4) \\ &\approx \cos(2\pi f_c t + \phi_3) + \frac{m}{2}\cos(2\pi f_{\text{upper}} t + \phi_3 + \phi_4 + \frac{\pi}{2}) + \frac{m}{2}\cos(2\pi f_{\text{lower}} t + \phi_3 - \phi_4 + \frac{\pi}{2}) \\ &\approx \cos(2\pi f_c t + \phi_3) + \frac{m}{2}\cos(2\pi f_{\text{upper}} t + \theta_{\text{upper}}) + \frac{m}{2}\cos(2\pi f_{\text{lower}} t + \theta_{\text{lower}}) \end{aligned}$$

Where we have set $\theta_{\text{upper}} = \phi_3 + \phi_4 + \frac{\pi}{2}$ and $\theta_{\text{lower}} = \phi_3 - \phi_4 + \frac{\pi}{2}$ giving,

$$\alpha_{PM} = (\theta_{\text{upper}} - \phi_3) - (\phi_3 - \theta_{\text{lower}}) = ((\phi_3 + \phi_4 + \frac{\pi}{2}) - \phi_3) - (\phi_3 - (\phi_3 - \phi_4 + \frac{\pi}{2})) = \pi,$$

concluding the mathematical derivation.

Experimentally, the encoding-type parameter α may take either of these values or any value between, and so a distribution of measured values is expected. α was calculated for all 9 sideband frequency pairs under the corresponding stimulus condition, for all 50 selected channels and all 11 subjects. Circular statistics were used to estimate the (circular) mean and (circular) standard error of α across all samples (Slow-FM Experiment: 600 samples =50 channels × 12 subjects; Fast-FM Experiment: 550 samples = 50 channels × 11 subjects) for each of the 9 corresponding sideband frequency pairs. To calculate the circular mean value $\bar{\alpha}$, for

each f_{FM}, all the α were first converted into complex vectors ($e^{i\alpha}$) and the mean of those complex vectors was determined. The circular mean $\bar{\alpha}$ is the four-quadrant inverse tangent of this complex vector mean. The circular standard error of α (SE_α) was calculated using bootstrap (balanced, 1000 instances) across the α of all samples. (Efron & Tibshirani, 1994; Fisher, 1996)

Vector strength of α

The vector strength of α (v_α, ranging between 0 and 1) is used to examine the robustness of the encoding-type parameter α. Larger v_α indicates narrower distribution of α, and smaller v_α indicates wider distribution of α (in fact $1-v_\alpha$ is mathematically equal to the circular variance of the distribution). It is defined as

$$v_\alpha = \frac{1}{N}\sqrt{(\sum_{i=1}^{N}\sin(\alpha_i))^2 + (\sum_{i=1}^{N}\cos(\alpha_i))^2}\ .$$

v_α is calculated for all 9 sideband frequency pairs and 9 stimulus conditions across all samples (Slow-FM Experiment: 600 samples; Fast-FM Experiment: 550 samples), giving a 9×9 matrix V_α. In V_α, the 9 rows represent the 9 sideband frequency pairs ($f_{AM} \pm f_{FM}$) and 9 columns represent the 9 different stimulus conditions. Each element in the matrix represents the v_α value of this specific sideband frequency pair (corresponding row) for a specific stimulus condition (corresponding column). Ideally, the corresponding stimulus condition should elicit a robustly narrow α distribution and therefore the maximum v_α value in each row.

Phase difference parameters

Another two phase parameters, $\theta_{Upperdiff}$ and $\theta_{Lowerdiff}$, are used to examine the phase properties of upper and lower sideband frequencies, respectively, complementary to the amplitude properties of sidebands characterized by A_{upper} and A_{lower}. They are defined as:

$$\theta_{Upperdiff} = \theta_{Upper} - \theta_{f_{AM}}$$
$$\theta_{Lowerdiff} = \theta_{Lower} - \theta_{f_{AM}}$$

Therefore,

$$\alpha = (\theta_{Upper} - \theta_{f_{AM}}) + (\theta_{Lower} - \theta_{f_{AM}}) = \theta_{Upperdiff} + \theta_{Lowerdiff}$$

$\theta_{Upperdiff}$ and $\theta_{Lowerdiff}$ were calculated for all target sideband frequencies (9 upper sideband frequencies and 9 lower sideband frequencies), all selected 50 channels, all 9 stimulus conditions, and all subjects (Slow-FM Experiment: 12 subjects; Fast-FM Experiment: 11 subjects). The same circular statistics used to calculate α were used to estimate the mean and standard error of $\theta_{Upperdiff}$ and $\theta_{Lowerdiff}$. Their vector strengths were defined as:

$$v_{Upperdiff} = \frac{1}{N}\sqrt{(\sum_{i=1}^{N}\sin\theta_{Upperdiff})^2 + (\sum_{i=1}^{N}\cos\theta_{Upperdiff})^2}$$
$$v_{Lowerdiff} = \frac{1}{N}\sqrt{(\sum_{i=1}^{N}\sin\theta_{Lowerdiff})^2 + (\sum_{i=1}^{N}\cos\theta_{Lowerdiff})^2}$$

These calculated vector strength values were used to construct two 9×9 vector strength matrices ($V_{Upperdiff}$, $V_{Lowerdiff}$) using the same configuration as that for V_α. In addition, $\theta_{Upperdiff}$ and $\theta_{Lowerdiff}$ were adjusted according to corresponding sideband frequencies to compensate for a group delay (latency) estimated by the slope of the $\theta_{Upperdiff}$-frequency and $\theta_{Lowerdiff}$-frequency curves, termed as $\theta^{adj}_{Upperdiff}$ and $\theta^{adj}_{Lowerdiff}$. Note

that they would have the same vector strength and standard errors as that of $\theta_{Upperdiff}$ and $\theta_{Lowerdiff}$.

Asymmetry index for amplitude and vector strength.

The amplitude asymmetry index AI_A and vector strength asymmetry index AI_V quantify any asymmetry between the upper and lower sidebands. They are normalized to lie between –1 and 1 and defined as

$$AI_A = \text{diag}\left(\frac{A_{Upper} - A_{Lower}}{A_{Upper} + A_{Lower}}\right)$$

$$AI_V = \text{diag}\left(\frac{V_{Upperdiff} - V_{Lowerdiff}}{V_{Upperdiff} + V_{Lowerdiff}}\right)$$

Simulations

We constructed a model neuron population whose SSR amplitude and phase are both modulated by the stimulus FM. In this model we posit that the phase modulation index is fixed (at $\pi/8$, as observed by Ross et al. (2001)), but the amplitude modulation index m may vary with f_{FM}. The goal was to see if an increase in the AM portion of the response, i.e. an increase in m, could account for the observed single sideband signal (SSB) for high f_{FM} rates. Simulated encoding signals with neural carrier frequency of 37 Hz (f_{AM}) and modulation frequency of 8 Hz (one example of f_{FM}) were created with additive Gaussian white noise (GWN) at a relative level of 15. The amplitude modulation index m varies from 0 to 0.8. The parameter θ, characterizing the phase shift of amplitude modulation contribution to $S(t)$ in relation

to the phase modulation contribution of $S(t)$, varies from 0 to 2π. We performed 300 simulations for each amplitude modulation index parameter *m* from 0 to 0.8 in step of 0.08 and for each phase shift parameter θ from 0 to 2π in step of $\frac{\pi}{4}$, and calculated parameters ($\theta^{adj}_{Upperdiff}$, $\theta^{adj}_{Lowerdiff}$, α, AI_A, AI_v) of the simulated signals.

$$S(t) = \underbrace{(1 + m\cos(2\pi f_{FM} t + \theta))}_{Amplitude\ Modulation} \times \underbrace{\cos(2\pi f_{AM} t + \frac{\pi}{8}\cos(2\pi f_{FM} t))}_{Phase\ Modulation} + GWN \qquad (1)$$

For comparison, we also constructed a model containing a pair of neural populations. The SSR *amplitude* of one population is modulated by the stimulus FM (AM encoding population, $S_{AM}(t)$), whereas the SSR *phase* of the other population is modulated by the stimulus FM (PM encoding population, $S_{PM}(t)$). Both $S_{AM}(t)$ and $S_{PM}(t)$ were created with carrier frequency of 37 Hz (f_{AM}) and modulation frequency of 8 Hz (one example of f_{FM}). The phase modulation index in $S_{PM}(t)$ and the amplitude modulation index in $S_{AM}(t)$ are fixed (at π/8 and 0.25 respectively, as measured by Ross et al. (2001)). The simulation mixed signal $S(t)$ were created by combining $S_{AM}(t)$ and $S_{PM}(t)$ using different mixing weights τ (pure PM: $\tau = 0$; pure AM: $\tau = 1$) and then by adding Gaussian white noise (GWN). The parameter θ, also characterizing the phase shift of amplitude modulation contribution to $S(t)$ in relation to the phase modulation contribution of $S(t)$, varies from 0 to 2π. The relative increase in the AM contribution to the response is given by the mixing weight parameter τ.

$$
\begin{aligned}
S_{PM}(t) &= \cos\left(2\pi f_{AM} t + \frac{\pi}{8}\cos(2\pi f_{FM} t)\right) \\
S_{AM}(t) &= \left(1 + 0.25\cos(2\pi f_{FM} t + \theta)\right)\cos(2\pi f_{AM} t) \\
S(t) &= \tau S_{AM}(t) + (1-\tau) S_{PM}(t) + GWN
\end{aligned}
\tag{2}
$$

Functionally, this paired neural population model is not distinguishable from the previous single neural population model since both of them test for the effect of an increase in the AM contribution of the response: an increase in *m* in the single neural population model and increase in τ in the paired neural population model). Either could account for the observed single sideband signal (SSB) for high f_{FM} rates. However, they are different in the hypothesized underlying neuron population structure and encoding properties.

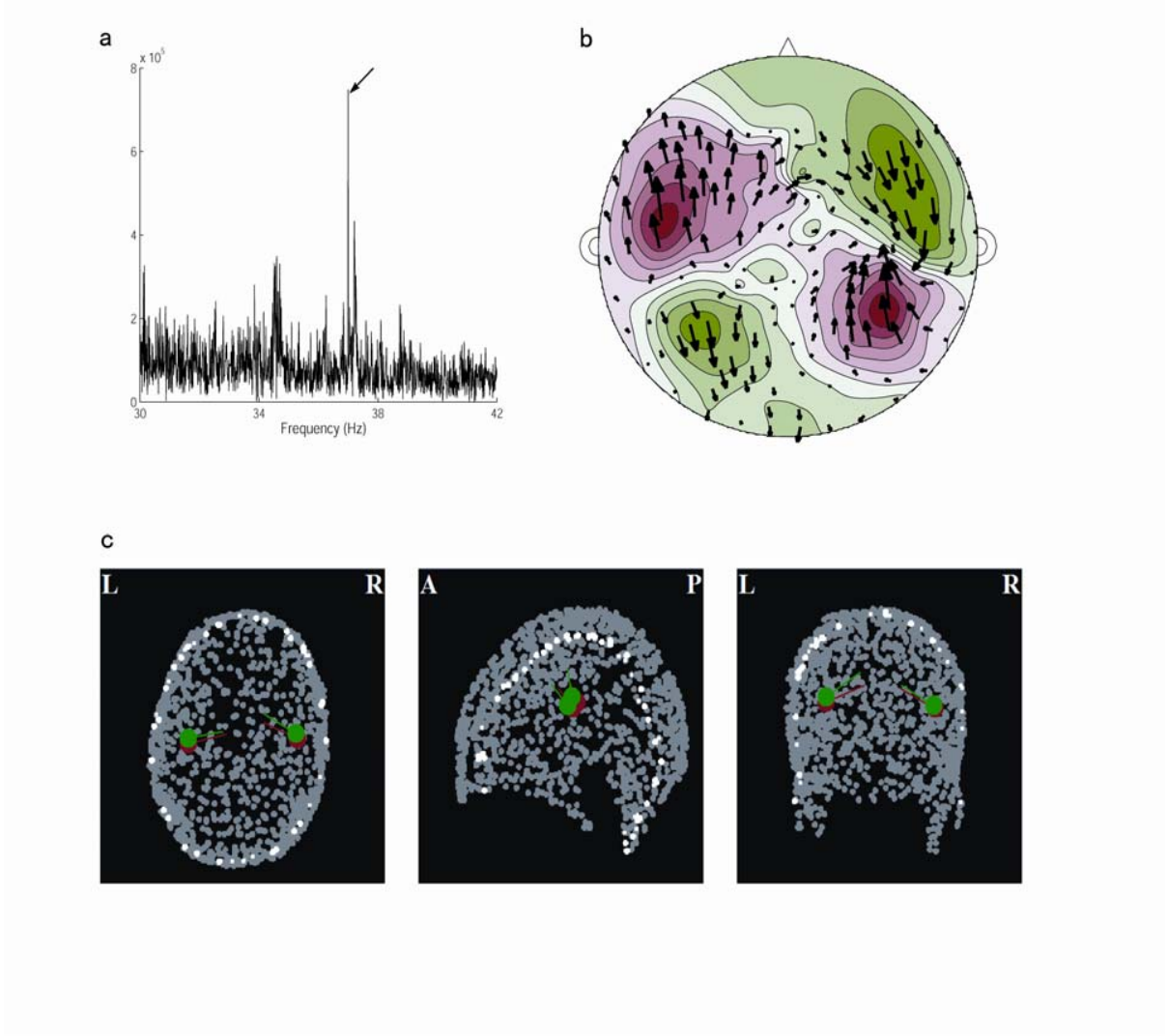
2.3 Results

2.3.1 Auditory steady-state response at f_{AM} and phasor representation

Clear stimulus-evoked aSSR at f_{AM} (37 Hz) was observed for all subjects under all 9 stimulus conditions in both Slow-FM Experiment and Fast-FM Experiment, since all the stimulus conditions have the same f_{AM} at 37 Hz and only differ in f_{FM} (Slow-FM Experiment: 0.3-8.0 Hz; Fast-FM Experiment: 2.1-30 Hz). Figure 2-4a shows the discrete Fourier transform of one channel of a representative subject in Slow-FM Experiment, including the aSSR at f_{AM} (37 Hz). The spectrum shows a clear peak at 37 Hz, the AM frequency f_{AM}. Because of the limited signal-to-noise ratio in the MEG signal, other peaks (external narrowband noise) are also observable (and known to be not due to movement or related artifacts, or from bad sensors). The

relevance of using sidebands to detect neural modulation coding is that the vast majority of the noise peaks cannot interfere with the sidebands. Figure 2-4b shows the corresponding phasor representations for aSSR at 37 Hz for all channels (Simon and Wang, 2005). There is a clear bilateral auditory cortical origin for aSSR at 37 Hz. Figure 2.4c shows the grand average results for both the aSSR equivalent-current dipole (red) and the M100 (green). The dipole locations of aSSR and of M100 activity were compared across all subjects, and it was found that they have displacement not significantly different from 0 (for right hemisphere: Δx = -1.1±5.3 mm, Δy = 4.6±7.6 mm, Δz = -2.4±5.8 mm; for left hemisphere: Δx = -0.0±3.2 mm, Δy = 4.4±8.2 mm, Δz = -4.1±5.4 mm). This result supports the idea that the source of aSSR is in superior temporal cortex since the M100 component is believed to originate there (Lutkenhoner and Steinstrater, 1998). This result is consistent with the aSSR localization results of Ross et al. (2000) given the resolution limitations of this data set. Figure 2-5 also illustrated the aSSR at f_{AM} at 37 Hz (Figure 2-5a: black arrows) in one representative subject in Fast-FM Experiment and corresponding phasor representation (Figure 2.5b).

Figure 2-4 Auditory steady state response (aSSR) at envelope modulation frequency (37 Hz). a) Spectrum of the response from one representative channel of one subject in Slow-FM Experiment. The arrow indicates the evoked aSSR at 37 Hz. b) The phasor representation of aSSR at 37 Hz. It clearly shows a bilateral auditory MEG contour map. The arrow in each channel represents the Fourier coefficient at 37 Hz. The arrow length represents the magnitude and the arrow direction represents the phase. c) Grand average of dipole location for the aSSR at 37 Hz (red) and M100 (green), in axial, sagittal and coronal views. The two dipoles are localized in similar position of superior temporal cortex.


a
x 10^5
8
6
4
2
0
30
34
38
42
Frequency (Hz)
b
c
L
R
A
P
L
R

2.3.2 Auditory steady-state response at sidebands

Figure 2-5 and Figure 2-6 show the aSSR at upper sidebands for the same channel in the same subject at different stimulus conditions for Slow-FM Experiment and Fast-FM Experiment respectively. First, the aSSR at 37 Hz (f_{AM}) can be seen for all 9 different stimulus conditions (black arrow); Secondly, stimuli with specific f_{FM} elicited corresponding sidebands (here, only upper sidebands are shown, grey arrows; the lower sidebands, not shown, do not necessarily follow the same pattern). For example, in Figure 2-5 (Slow-FM Experiment), for stimulus f_{FM} = 0.5 Hz, the response at 37.5 Hz (= 37 + 0.5) is elicited, and when stimulus f_{FM} = 1 Hz, the response at 38 Hz (= 37 + 1) is elicited. For this one channel, the upper sideband for f_{FM} of 5 Hz is not visible. Note that narrowband noise coexists with the sidebands we want to detect. Similarly, as illustrated in Figure 2-6 (Fast-FM Experiment), each stimulus with different f_{FM} (2.1–30 Hz) elicited corresponding sidebands, indicated by grey arrows. For example, for stimuli with f_{FM} of 8 Hz, the response spectrum showed a peak at 45 Hz (37 + 8 Hz); for stimuli with f_{FM} of 10.3 Hz, there was a peak at 47.3 Hz (37 + 10.3 Hz); for stimuli with f_{AM} of 15.1 Hz, a spectral peak at 52.1 Hz (37 + 15.1 Hz) was elicited. Note that Figure 2-5 and Figure 2-6 both illustrate the spectrum of the same channel under 9 different stimulus conditions, and it clearly indicates that in this case the observed sideband frequency peak was stimulus-elicited. Direct FM generated aSSR, i.e. at the frequencies of f_{FM} (2.1–30 Hz), were also observed, in agreement with previous findings (Picton et al., 1987; Dimitrijevic et al., 2001; Luo et al., 2006). For example, the stimulus with f_{FM} of 8 Hz elicited an aSSR response peak at 8 Hz, and the stimulus with f_{FM} of 30 Hz

elicited an aSSR response peak at 30 Hz, in addition to any sidebands around the AM generated SSR.

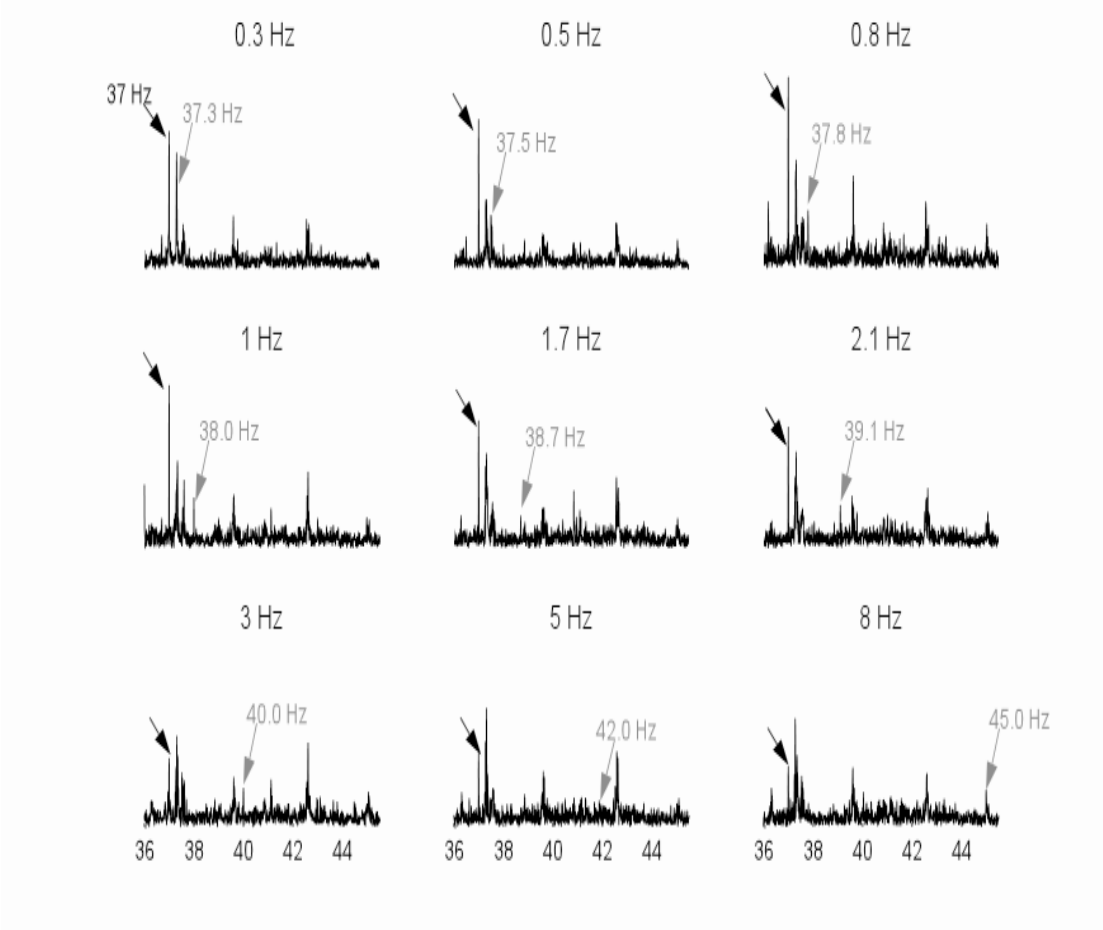


Figure 2-5 Spectrum and auditory-steady state response (aSSR) at sidebands at one channel in a representative subject in Slow-FM Experiment. Each of the 9 figures represents the spectrum for each of the 9 different f_{FM} stimulus conditions. The black arrow points to the aSSR at envelope modulation frequency (37 Hz) and can be observed for all the stimulus conditions. The grey arrows indicate the aSSR at corresponding upper sideband ($f_{AM} + f_{FM}$). For example, the stimulus with f_{FM} of 0.3 Hz elicited 37.3 Hz aSSR (grey arrow). For this specific channel, all the stimulus conditions elicited corresponding upper sidebands except the stimulus with f_{FM} of 5 Hz (grey arrow).

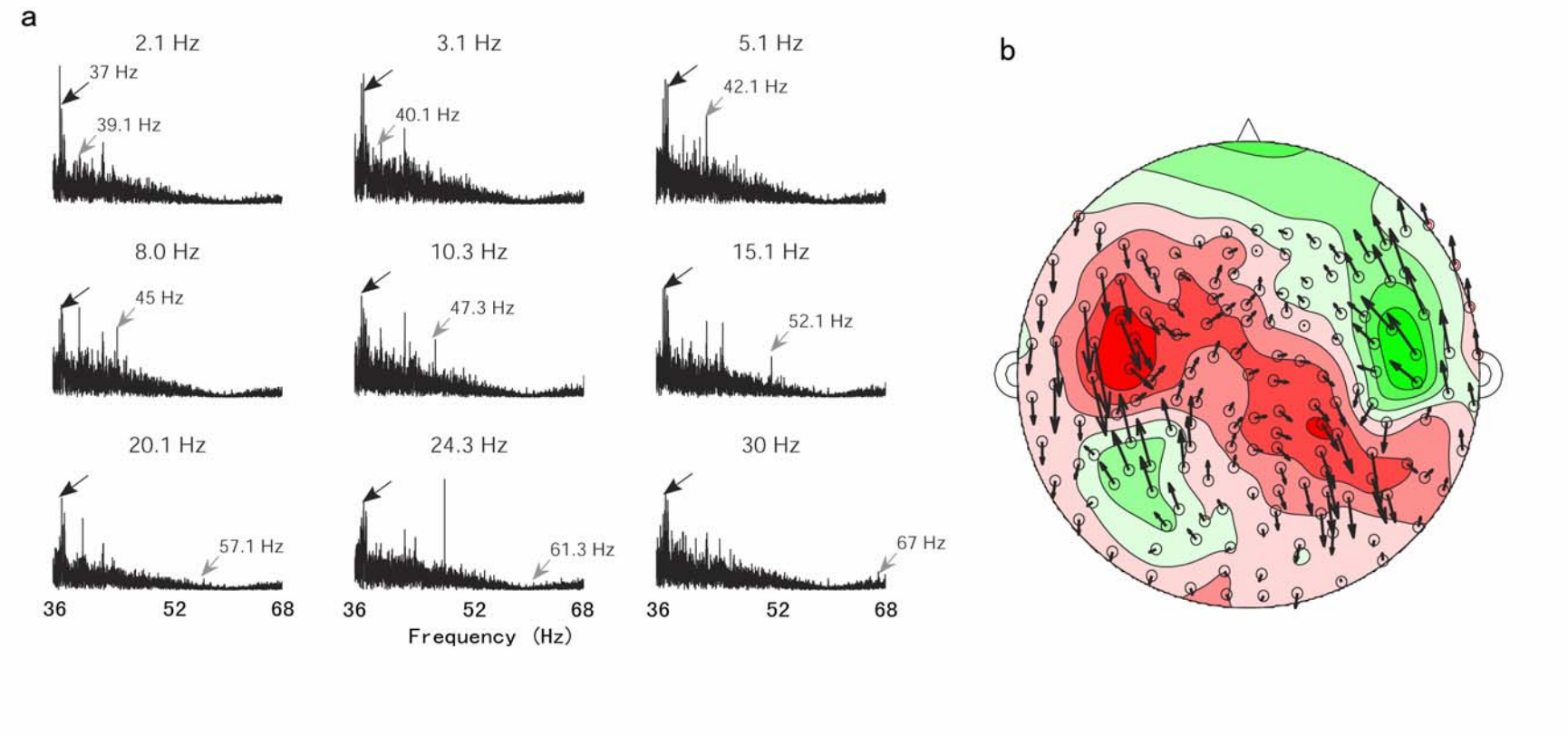

a
2.1 Hz
37 Hz
39.1 Hz
3.1 Hz
40.1 Hz
5.1 Hz
42.1 Hz
8.0 Hz
45 Hz
10.3 Hz
47.3 Hz
15.1 Hz
52.1 Hz
20.1 Hz
57.1 Hz
24.3 Hz
61.3 Hz
30 Hz
67 Hz
36
52
68
Frequency (Hz)
b

Figure 2-6 Auditory steady state response (aSSR) at f_{AM} (37 Hz) and upper sideband (37 + f_{FM}). a) Spectrum of the response from one representative channel of one subject under all 9 stimulus conditions (different stimulus f_{FM}), denoted by the subtitle value. The black arrows point to the aSSR at f_{AM} (37 Hz) and the grey arrows indicate the corresponding upper sideband frequency (f_{AM} + f_{FM}) for each specific stimulus condition. For example, the stimulus with f_{FM} of 8.0 Hz (the first figure in the second row) elicited aSSR at 45 Hz (37+8.0, gray arrow). b) The phasor representation of aSSR at f_{AM} (37 Hz). It clearly shows a bilateral auditory MEG contour map. The arrow length in each channel indicates the aSSR amplitude at 37 Hz, and the arrow direction represents the aSSR phase. Note that the channels with largest arrows (largest aSSR at 37 Hz) are centered in the bilateral auditory cortex positions, representing the origin of the elicited aSSR and are the main places where the 50 channels were selected from for further analysis.

2.3.3 Transition from two sidebands to one sideband

As can be seen in Figure 2-5 and Figure 2-6, narrow-band system noises coexist with the spectral responses we want to detect (f_{AM}, f_{FM}, sidebands), which in turn makes the direct detection of the narrowband response at sideband frequencies more difficult. A method to check the significance of the narrowband response, elicited at a target sideband frequency by the corresponding stimulus, is by an across-condition comparison, shown in the sideband amplitude matrices (A_{Upper}, A_{Lower}).

Figure 2-7 shows the grand average of the upper (A_{Upper}) and lower (A_{Lower}) sideband amplitude matrices across subjects, and for both Experiment I (Figure 2-7a,d; f_{FM}: 0.3–8 Hz; 12 subjects) and Fast-FM Experiment (Figure 2-7b, e; f_{FM}: 2.1–30 Hz; 11 subjects). In these matrices, most rows peak on the diagonal, which indicates that those sideband frequencies (rows) were significantly elicited by the

corresponding stimulus (and not by any other stimulus condition). In addition, there are noticeable differences between A_{Upper} (Figure 2-7a,b) and A_{Lower} (Figure 2-7d,e). Specifically, A_{Upper} shows a strong dominantly diagonal pattern, whereas this pattern was much murkier and noisier in A_{Lower}, especially in the high f_{FM} range (Figure 2-7e).

Such asymmetrical performance between A_{Upper} and A_{Lower} can be seen more clearly in plots below in Figures 2-7c and 2-7f, illustrating the corresponding 9-value diagonal value vector of A_{upper} and A_{lower}, respectively. Slow-FM Experiment data (grey line) and Fast-FM Experiment results (black line) are plotted in the same figure for comparison. Note that there are 4 overlapping f_{FM} stimulus conditions (2.1, 3.1, 5.1, 8 Hz) which show consistently good results. The horizontal starred line indicates the mean amplitude level at this frequency, i.e. the noise floor. Specifically, for stimuli with low f_{FM} ($\lesssim$ 5 Hz), both upper and lower sidebands are strongly elicited (with the exception of two outliers in the upper sideband, f_{FM} at 0.3 and 0.5 Hz, which are artificially small due to system narrowband noise at 37.3 and 37.5 Hz (Luo et al., 2006)). For stimuli with faster FM rates (5 Hz $\lesssim$ f_{FM} $\lesssim$ 24 Hz), there is an asymmetry between the upper and lower sideband amplitudes: as the modulation frequency increases, the lower sideband level decreases toward the noise floor, whereas the upper sideband level remains well above the noise floor. For the fastest stimuli (f_{FM} $\gtrsim$ 24 Hz), both upper and lower sidebands decrease to the noise floor. In summary, we observed a two-sideband-to-one-sideband spectral pattern transition with increasing stimulus f_{FM} (up to 24.3 Hz), with the transition occurring at f_{FM} ~ 5 Hz.

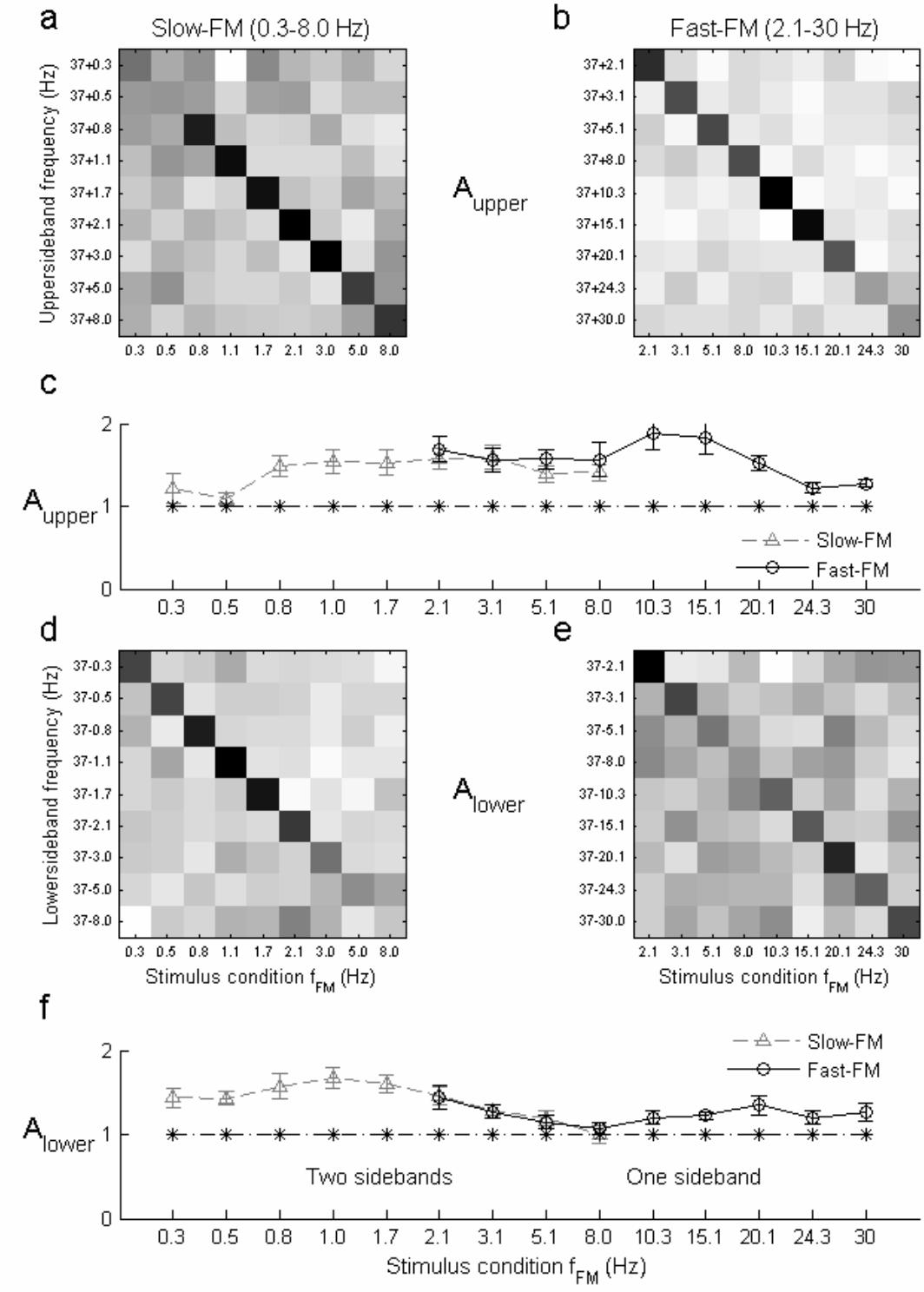

a
Slow-FM (0.3-8.0 Hz)
Uppersideband frequency (Hz)
37+0.3
37+0.5
37+0.8
37+1.1
37+1.7
37+2.1
37+3.0
37+5.0
37+8.0
0.3 0.5 0.8 1.1 1.7 2.1 3.0 5.0 8.0
A upper
b
Fast-FM (2.1-30 Hz)
37+2.1
37+3.1
37+5.1
37+8.0
37+10.3
37+15.1
37+20.1
37+24.3
37+30.0
2.1 3.1 5.1 8.0 10.3 15.1 20.1 24.3 30
c
A upper
2
1
0
Slow-FM
Fast-FM
0.3 0.5 0.8 1.0 1.7 2.1 3.1 5.1 8.0 10.3 15.1 20.1 24.3 30
d
Lowersideband frequency (Hz)
37-0.3
37-0.5
37-0.8
37-1.1
37-1.7
37-2.1
37-3.0
37-5.0
37-8.0
0.3 0.5 0.8 1.1 1.7 2.1 3.0 5.0 8.0
Stimulus condition f FM (Hz)
A lower
e
37-2.1
37-3.1
37-5.1
37-8.0
37-10.3
37-15.1
37-20.1
37-24.3
37-30.0
2.1 3.1 5.1 8.0 10.3 15.1 20.1 24.3 30
Stimulus condition f FM (Hz)
f
A lower
2
1
0
Slow-FM
Fast-FM
Two sidebands
One sideband
0.3 0.5 0.8 1.0 1.7 2.1 3.1 5.1 8.0 10.3 15.1 20.1 24.3 30
Stimulus condition f FM (Hz)

Figure 2-7 Amplitude matrix A_{Upper} and A_{lower} for both Slow-FM Experiment (f_{FM}: 0.3–8 Hz) and Fast-FM Experiment (f_{FM}: 2.1–30 Hz), and the corresponding diagonal value vectors. a) A_{Upper} of Slow-FM Experiment. b) A_{Upper} of Fast-FM Experiment. d) A_{Lower} of Slow-FM Experiment. e) A_{Lower} of Fast-FM Experiment. Each box represents the normalized amplitude at the particular target upper sideband frequency (vertical axis) under specific stimulus condition (horizontal axis). c) Diagonal value vectors of A_{Upper} (grey dotted line: Slow-FM Experiment; black solid line: Fast-FM Experiment). f) Diagonal value vectors of A_{Lower} for both Slow-FM Experiment and Fast-FM Experiment. The starred lines in the plots indicate the noise floor at each specific target frequency. Note that f_{FM} around 5.0 Hz marked the transition from 'Two sidebands' to 'One sideband'.

2.3.4 Transition from PM to unreliable encoding-type parameter α

Sidebands naturally occur for all types of modulation coding (including AM and PM). To help determine which modulation coding created the sidebands, an encoding-type parameter (α, defined below, ranging between 0 and 2π) was calculated to distinguish AM encoding from PM encoding. Both encoding mechanisms (see Figure 2-1) elicit two sidebands, but with different phase relationships across the sidebands and carrier.

Figure 2-8 summarizes the behavior of the encoding-type parameter α for both Slow-FM Experiment (grey line) and Fast-FM Experiment (black line). Figure 2-8a shows the circular mean and standard error of α, which lies roughly in the PM encoding region (~ π) for slower f_{FM} ($\lesssim$ 5 Hz) and transitions into a regime of undetermined values with increasing f_{FM} rate (as stated above, the outlier at f_{FM} = 0.3 Hz is due to the narrowband system noise at 37.3 Hz).

The vector strength of α (V_α) is calculated to further examine the robustness of modulation encoding and the reliability of observed modulation encoding type. Figure 2-8b illustrates the entire matrix V_α and the corresponding diagonal value vectors for both Slow-FM Experiment and Fast-FM Experiment. Specifically, the V_α for Slow-FM Experiment (f_{FM}: 0.3–8 Hz, left matrix of Figure 2-8b) manifests a dominantly diagonal pattern, especially for f_{FM} below 5 Hz (1st to 7th row), compared to the V_α for Fast-FM Experiment (f_{FM}: 2.1–30 Hz, right matrix of Figure 2-8b), in which only the first 3 rows (corresponding to f_{FM} of 2.1, 3.1, and 5.1 Hz) show a dominant diagonal. The corresponding diagonal value vectors (Figure 2-8c) decrease to the baseline v_α value as stimulus f_{FM} increases, reflecting that the encoding-type parameter α becomes increasingly noisier and more unreliable for faster stimulus f_{FM} ($\gtrsim$ 5 Hz), although the α value seemed to shift roughly to the AM encoding region (~ 0 or 2π) in Figure 4a. In summary, we observe a transition of α from the PM encoding region (~ π) to unreliable and noisy values as the stimulus rate f_{FM} increases, with a transition point of f_{FM} ~ 5 Hz.

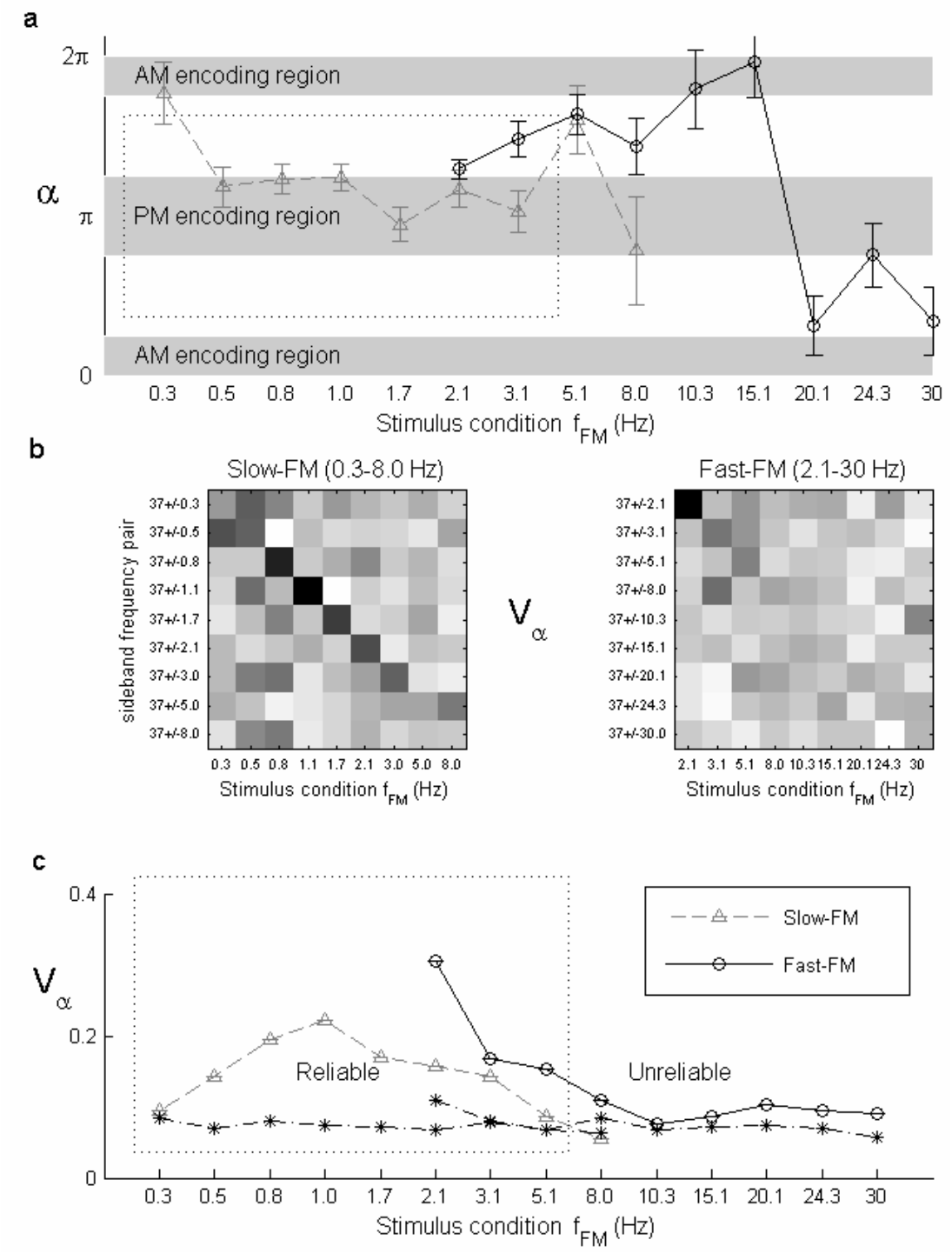


Figure 2-8 Encoding-type parameter α and the corresponding vector strength matrix V_α. a) α plot for different f_{FM} stimulus conditions (grey dotted line: Slow-FM Experiment; black solid line: Fast-FM Experiment) using circular statistics. Grey bars represent the PM encoding region

(middle, ~ π) and AM region (upper and lower, ~ 0 or 2π). b) V_α of both Slow-FM Experiment (left matrix) and Fast-FM Experiment (right matrix). Each box in the matrix represents the vector strength calculated from α samples (Slow-FM Experiment: 600 samples; Fast-FM Experiment: 550 samples), for the specific sideband frequency pair (row) under different stimulus conditions (column). Vector strength is also equal to 1 minus the circular variance of the distribution. c) Diagonal vectors of V_α (Slow-FM Experiment: grey dotted line; Fast-FM Experiment: black solid line). The starred line indicates the corresponding mean vector strength value of each row representing the background vector strength of α across all 9 stimulus conditions. Note that f_{FM} around 5.0 Hz again marked the transition from 'Reliable' to 'Unreliable'. The dotted rectangle indicates the reliable range of α, where corresponding v_α is above background level (starred line).

2.3.5 Transition from symmetry to asymmetry in phase

There are two primary motivations to investigate the behavior of $\theta_{Upperdiff}$ and $\theta_{Lowerdiff}$, the two subcomponents of α. First, as stated above, we found that α becomes noisier and unreliable for $f_{FM} \gtrsim 5$ Hz (Figure 2-8), but at least the upper sideband response is still significantly elicited (Figure 2-7a, b), indicating the sustained presence of some form of modulation encoding. Therefore, by examining the corresponding changes of these two subcomponents of α, we can show underlying reasons for α becoming noisier. Secondly, we can use these subcomponents to investigate the phase performances for the upper and lower sidebands separately, as we did for amplitude analysis. From the signal processing side, the vector strengths of $\theta_{Upperdiff}$ and $\theta_{Lowerdiff}$ ($v_{Upperdiff}$, $v_{Lowerdiff}$) reflect the temporal precision (latency, starting phase, etc.) of the elicited MEG response.

Figure 2-9 illustrates the $V_{Upperdiff}$ (a-c) and $V_{Lowerdiff}$ (d-f) results for both Slow-FM Experiment (Figure 2-9a, d) and Fast-FM Experiment (Figure 2-9b, e). We can observe the dominantly diagonal pattern in most of the four matrices, indicating that these phase parameters ($\theta_{Upperdiff}$, $\theta_{Lowerdiff}$) manifested smaller variance (larger vector strength) under the corresponding stimulus conditions (compared to other stimulus conditions). In addition, there is some difference between $V_{Upperdiff}$ (Figure 2-9a, b) and $V_{Lowerdiff}$ (Figure 2-9d, e). Specifically, $V_{Upperdiff}$ showed a dominantly diagonal pattern, whereas this pattern is much murkier and noisier in $V_{Lowerdiff}$, especially for the high f_{FM} range (Figure 2-9e). Such asymmetrical behavior between $V_{Upperdiff}$ and $V_{Lowerdiff}$ is also reflected in Figures 2-9c and 2-9f, which illustrate the corresponding 9-value diagonal value vector of $V_{upperdiff}$ and $V_{lowerdiff}$,respectively, for both Slow-FM Experiment (grey line) and Fast-FM Experiment (black line). The horizontal starred line indicates the mean vector strength for this phase parameter across all stimulus conditions. Specifically, for stimuli with low f_{FM} ($\lesssim$ 5 Hz), vector strengths for both $\theta_{Upperdiff}$ and $\theta_{Lowerdiff}$ ($v_{Upperdiff}$, $v_{Lowerdiff}$) were significantly above the noise floor (with the exception of the f_{FM} = 0.3 Hz outlier in $v_{upperdiff}$, due to system narrowband noise at 37.3). For stimuli with faster FM ($f_{FM} \gtrsim$ 5 Hz), there is an asymmetry between $v_{Upperdiff}$ and $v_{Lowerdiff}$: $v_{Lowerdiff}$ decreases toward the noise floor (Figure 2-9a-c) whereas $v_{Upperdiff}$ remains well above (Figure 2-9d-f). In summary, with increases in stimulus f_{FM}, we observe a symmetry-to-asymmetry transition in the vector strength of phase parameters between upper and lower sidebands, where the transition point is f_{FM} ~ 5 Hz. This symmetry-to-asymmetry transition is similar to the two-to-one sideband transition in the amplitude matrix (Figure 2-7), indicating a certain

relationship between the two groups of parameters: the phase parameters ($V_{Upperdiff}$, $V_{Lowerdiff}$) and the amplitude parameters (A_{Upper}, A_{Lower}).

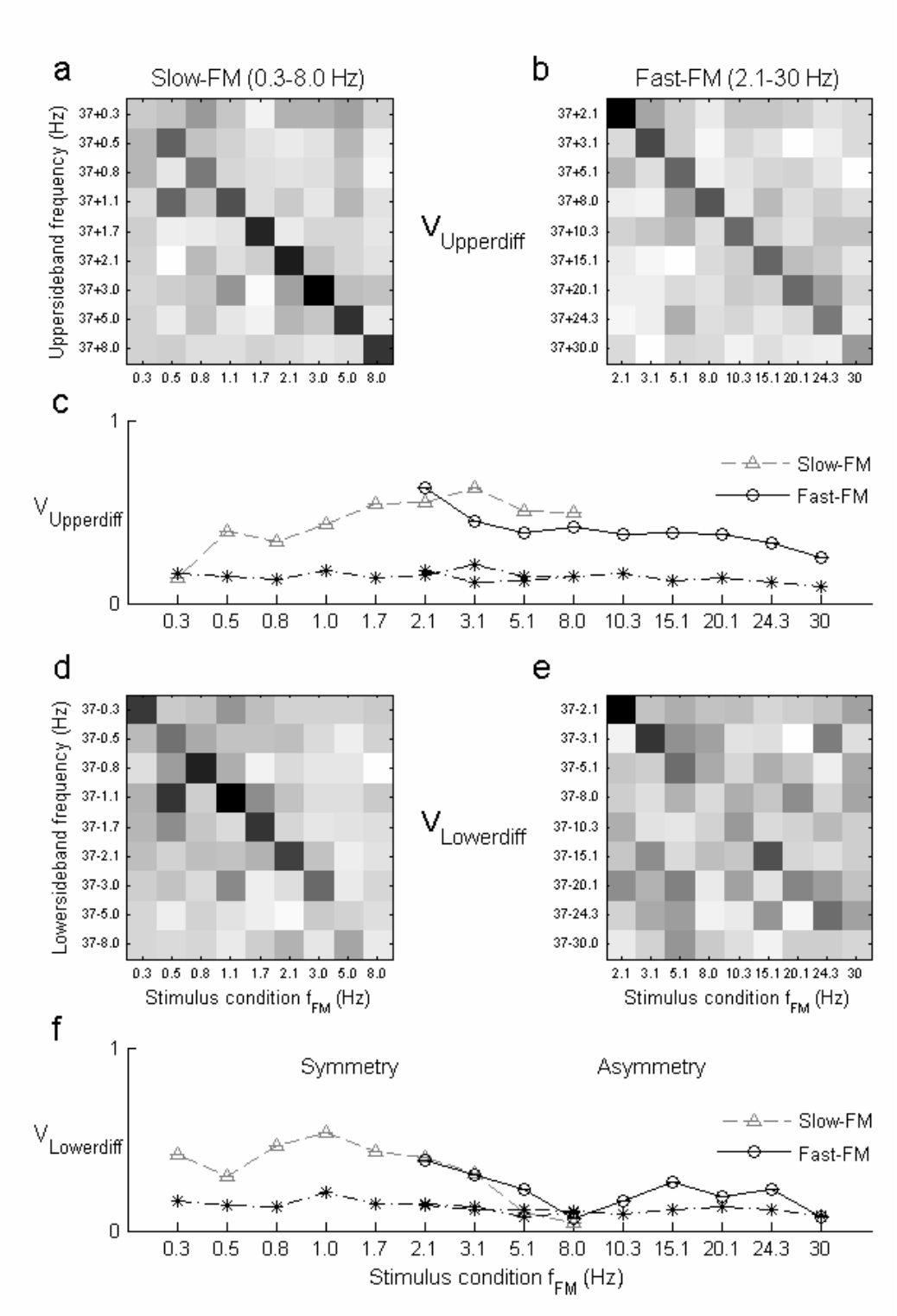


Figure 2-9 Phase vector strength matrix $V_{Upperdiff}$ and $V_{lowerdiff}$ for both Slow-FM Experiment (f_{FM}: 0.3–8 Hz) and Fast-FM Experiment (f_{FM}: 2.1–30 Hz), and the corresponding diagonal value vectors. a) $V_{Upperdiff}$ of Slow-FM Experiment. b) $V_{Upperdiff}$ of Fast-FM Experiment. c) $V_{Lowerdiff}$ of Slow-FM Experiment. d) $V_{Lowerdiff}$ of Fast-FM Experiment. Each box represents the calculated vector strength of the specific phase parameter ($\theta_{Upperdiff}$, $\theta_{Lowerdiff}$) (vertical axis) under specific stimulus condition (horizontal axis). c) Diagonal value vectors of $V_{Upperdiff}$ (grey dotted line: Slow-FM Experiment; black solid line: Fast-FM Experiment). f) Diagonal value vectors of $V_{Lowerdiff}$ for both Slow-FM Experiment and Fast-FM Experiment. The starred lines indicate the mean of each corresponding row, indicating the phase vector strength background level. Note that f_{FM} around 5.0 Hz marked the transition from 'Symmetry to 'Asymmetry'.

Additionally, $\theta_{Upperdiff}$ and $\theta_{Lowerdiff}$ were adjusted to compensate for a 40 ms group delay (latency) estimated from the $\theta_{Upperdiff}$-frequency and $\theta_{Lowerdiff}$-frequency curves. This 40 ms value also matches well with the results of Ross et al. (2000). The circular means and standard errors of the adjusted $\theta_{Upperdiff}$ and $\theta_{Lowerdiff}$ are plotted in Figure 2-10a.

2.3.6 Transition in both amplitude and phase from symmetry to asymmetry

The strong correlation between the phase and amplitude parameters for both Slow-FM Experiment (triangle) and Fast-FM Experiment (circle) is summarized in Figure 2-10c&d, which plot the amplitude asymmetry index AI_A and the phase vector strength asymmetry index AI_V, respectively, as a function of f_{FM} . Specifically, both AI_A and AI_V are near zero for the lowest and highest f_{FM} ranges (f_{FM} < 5.1 Hz, f_{FM} > 20.1 Hz), indicating commensurate results in both amplitude and phase reliability between the upper and lower sidebands (as before, the two outliers at f_{FM} of 0.3 and

0.5 Hz are due to system narrowband noise at 37.3Hz and 37.5 Hz). For the middle f_{FM} range (5 Hz $\lesssim$ f_{FM} $\lesssim$ 20 Hz), both AI_A and AI_V increase significantly above zero, indicating the emergence of an asymmetry between the upper and lower sideband responses; here the asymmetry favors the upper sideband in both amplitude and phase. These results are consistent with the previous amplitude, encoding-type parameter, and phase results (Figure 2-7, 2-8, 2-9), and they reconfirm the coding transition from pure PM encoding (two elicited sidebands, robust phase at both sidebands, α approximately π) to a different encoding strategy (elicited upper sideband only, robust phase at only upper sideband, α becoming noisier and unreliable). In summary, a transition from PM encoding to single sideband encoding (SSB) is confirmed here.

2.3.7 Simulation results

Figure 2-10 shows simulation results for a single neural population model. The simulation results, a function of both modulation index m and phase shift parameter θ are illustrated in Figure 2-10i-l, in matrix form. For each θ, all the simulations show complex transitions as m changes. The results for $\theta = \frac{\pi}{2}$, shown in Figure 2-10e-h, show transitions that are strikingly similar to those found in the data (Figure 2-10a-d), not only for the measured parameters (Figure 2-10a,b) but also for their distributions (Figure 2-10c,d). For other values of θ, the matching performance may be good for some of the parameters, but not all of them. These results suggest that introduction of fixed 90-degree phase delay, a quadrature relationship, between the amplitude

modulation contribution to *S(t)* and the phase modulation contribution to *S(t)* is necessary to account for the observed PM-to-SSB transition as we observed.

The simulation results for $\theta = \frac{\pi}{2}$ (Figure 2-10e-h) can be divided into 3 regions: PM-dominated, PM-AM-mixture, and AM-dominated, corresponding to small, middle, and large *m*, respectively. The most interesting and relevant range is the mixture region. Specifically, as the role of the subsidiary AM encoding increases (increasing *m*), $\theta^{adj}_{Upperdiff}$ (black line) remains relatively fixed with small error bars throughout the range of *m*, whereas $\theta^{adj}_{Lowerdiff}$ (grey line) manifests a rough transition through π and with larger error bars. At the same time, the encoding type parameter α shows a transition from PM encoding region (~π) to AM encoding region (~0). As for the asymmetry index performance, both AI_A and AI_V are strongly positive in the mixture range, reflecting the dominance of the upper sideband in the signals. The simulation results match the empirical results in many facets (Figure 2-10a-d), suggesting that the observed transition from a PM encoding signal to a SSB signal may be due to the increasing importance of a subsidiary AM encoding mechanism (invoked in the simulation by increasing the amplitude modulation index *m*) in addition to the already present PM encoding, as a monotonic function of f_{FM}.

Similar simulation results are found from the paired neural population model and thus not illustrated here. Specifically, the simulation results can also be divided into three regions: PM-dominated, PM-AM-mixture, and AM-dominated, corresponding

to small, middle, and large τ, respectively. As τ increases, $\theta^{adj}_{Upperdiff}$, $\theta^{adj}_{Lowerdiff}$, α, AI_A, and AI_V of simulated signals showed the similar transition pattern as that of single neural population model in Figure 2-10, suggesting that additional involvement of activities of a subsidiary AM encoding neural population in the response of an already present PM encoding neural population could account for the observed transition from pure PM encoding to SSB signal. Importantly, it also requires a 90-degree phase shift between the two neural populations' modulation signals ($S_{AM}(t)$ and $S_{PM}(t)$). Such a precise phase relationship between two independent neural populations is an extra required assumption, which leads us to favor and emphasize the single neural population model.

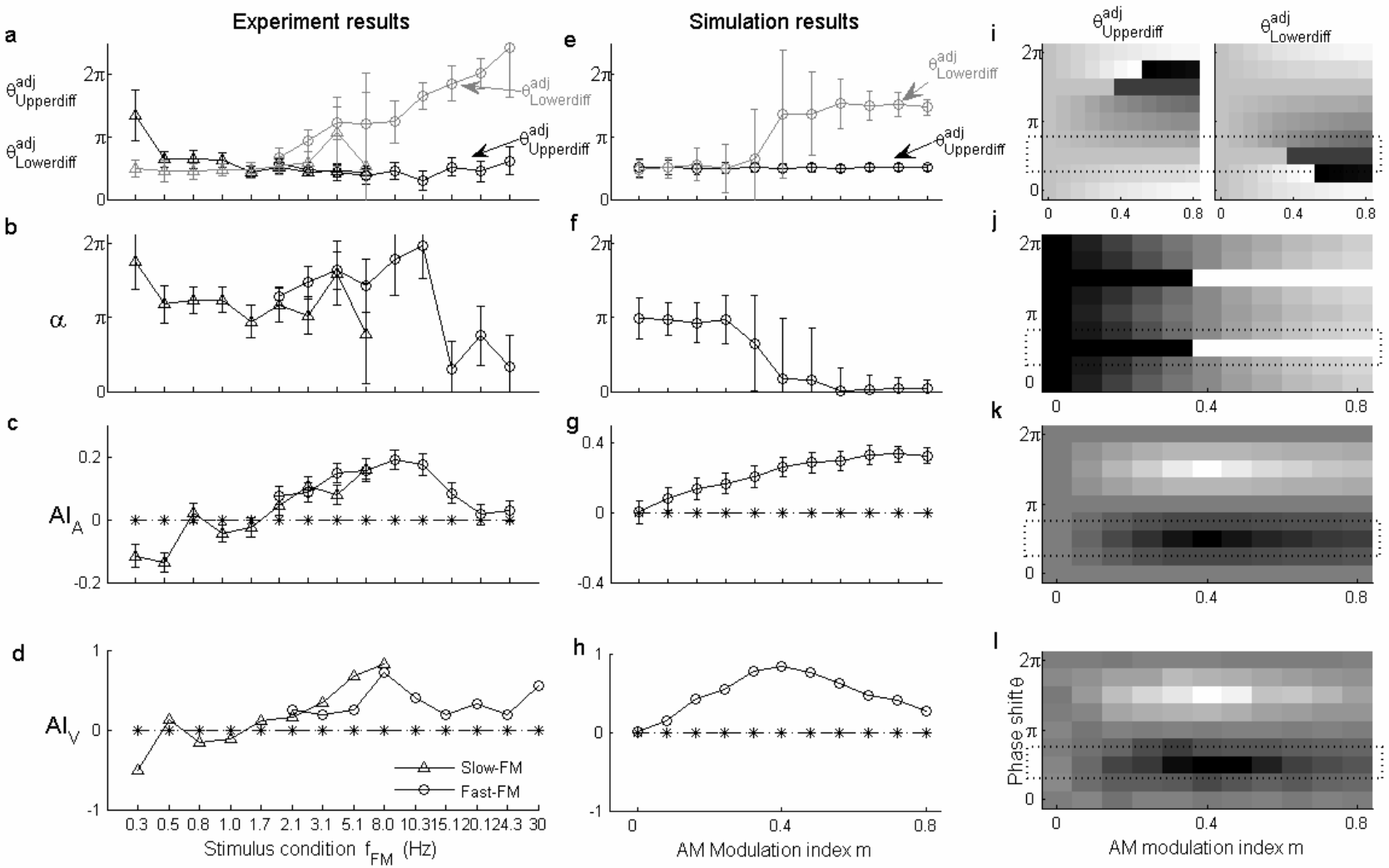

Experiment results
Simulation results
a
b
c
d
e
f
g
h
i
j
k
l
$\theta^{adj}_{Upperdiff}$
$\theta^{adj}_{Lowerdiff}$
α
AI_A
AI_V
Slow-FM
Fast-FM
0.3 0.5 0.8 1.0 1.7 2.1 3.1 5.1 8.0 10.315.120.124.3 30
Stimulus condition f_{FM} (Hz)
AM Modulation index m
Phase shift θ
AM modulation index m

Figure 2-10 Comparisons between experiment results (a–d, Slow-FM Experiment: triangle; Fast-FM Experiment: circle) and simulation results (i–l: simulation matrix results as a function of both AM modulation index *m* and phase shift parameter *θ*; e–h: simulation result plots for *θ* at $\frac{\pi}{2}$). a,e,i) $\theta^{adj}_{Upperdiff}$ (black line) and $\theta^{adj}_{Lowerdiff}$ (grey line). b,f,j) Encoding type parameter *α*. c,g,k) Amplitude asymmetry index (AI_A) between upper and lower sideband. d,h,l) Phase vector strength asymmetry index (AI_V) between phase parameter $\theta^{adj}_{Upperdiff}$ and $\theta^{adj}_{Lowerdiff}$. The starred line at 0 in AI_A and AI_V plots indicate the symmetrical performance between upper and lower sideband performances. Black boxes indicate large values in matrices results. Note that the simulation result plots (e–h) reproduce to one part of (blue rectangle) the simulation matrix results (right column). Note that the simulation results for *θ* at $\frac{\pi}{2}$ (e-h) matched well with the experiment results (a-d): $\theta^{adj}_{Upperdiff}$ (a,e, black line) remains flat with small error bars; $\theta^{adj}_{Lowerdiff}$ (a,e, grey line) manifests a rough transition through π with larger error bars; encoding type parameter α (b,f) shows a transition from ~π to ~0; both AI_A and AI_V manifest a transition from 0 to positive values (c,d,g,h).

2.4 Discussion

In this set of experiments, we have investigated the mechanisms of co-representation of simultaneous acoustic AM and FM, two of the most significant acoustic properties of natural communication sounds, for a possible coding transition at increased stimulus dynamic rates. Using sounds with simultaneous sinusoidally modulated amplitude (AM, f_{AM} = 37 Hz) and carrier frequencies (FM, f_{FM} = 0.3–30 Hz), the elicited MEG responses were analyzed. We had two important findings:

First, by confirming the presence of elicited spectral sideband pattern, **modulation encoding** was found in human auditory cortex to co-represent the envelope and carrier dynamics simultaneously. Secondly, we observed a **modulation encoding transition** in the MEG responses with increase of stimulus dynamic rates, from pure PM encoding signals, to signals containing only the upper sideband in the spectrum (SSB). A neuronal model was constructed and suggested that the introduction of a subsidiary AM encoding mechanism onto the already present PM encoding would explain the occurrence of SSB encoding.

2.4.1 Relationship to previous aSSR findings

Consistent with previous research (Ross et al., 2000), we find a robust aSSR at f_{AM} (37 Hz here), which means auditory cortex demodulates the incoming sound and extracts the envelope. The aSSR at f_{FM} is consistent with EEG studies using pure frequency modulated stimuli (Picton et al., 1987), which is one way auditory cortex represents pure carrier dynamics, although they tested much higher modulation frequencies (>80 Hz) than those used here. Dimitrijevic et al. (2001), used independent amplitude and frequency modulation (IAFM) stimuli with also higher modulation frequencies and found separate AM and FM aSSR responses that are relatively independent of each other, suggesting separate and independent encoding of envelope and carrier. We also found the aSSR at f_{FM}, but since our AM frequency was fixed, we cannot estimate whether the aSSR at f_{AM} and f_{FM} were independent of each other. When the source of the aSSR was localized using equivalent-current

dipoles, no significant difference was found between the location of these dipoles and those of the (well-studied) M100.

We found that for slower f_{FM} stimuli (< 5 Hz), the encoding-type parameter α is approximately π, indicating that those sidebands are due to the *phase* modulation of f_{AM} by f_{FM}. In other words, the phase of the aSSR at f_{AM} tracked the stimulus carrier frequency change, and because the carrier frequencies changed at certain frequencies (f_{FM}), the phase of f_{AM} also changed at the corresponding f_{FM} frequencies. These results for slower f_{FM} were consistent with Patel & Balaban (2000) where the phase of the aSSR reliably tracked the carrier frequency contour of the tone sequences. There the carrier was a long, periodic, series of concatenated tone segments (f_{FM} ~0.1 Hz), rather than the sinusoidally modulated carrier in our experiment. These results suggest that for stimuli with slow carrier dynamics (f_{FM} < 5 Hz), auditory cortex tracks the carrier dynamics, i.e. the stimulus carrier frequency change, by modulating the phase of the aSSR at f_{AM} accordingly.

2.4.2 Modulation encoding for feature grouping

Temporal modulation features characterize the dynamics in a sound. AM describes changes in temporal amplitude (envelope), and FM describes changes in carrier frequency (fine structure). Stimuli with temporal modulation features are often used to examine the extent to which sensory neurons can fire spikes following the temporal structures of the stimuli. For this reason the concept of modulation is useful to describe both the stimulus dynamics and the corresponding stimulus-locked

responses. Elhilali et al. (2004) have shown that in ferret AI, neural responses lock not only to envelope dynamics (e.g. AM), but also to the carrier dynamics (e.g. FM). Cariani (2004), among the many possible temporal neural codes, proposes multiplexing, a method widely used in telecommunication, as a perceptual grouping mechanism. In this view, the same neural element may be responsible for both concurrent representation and transmission of multiple signals. Similarly here, by observing a significant spectral peak with robust phase behavior at sideband frequencies, we demonstrate that modulation encoding, an efficient encoding method to multiplex two features' representations, can track stimulus AM and FM simultaneously. It provides a natural means of perceptual grouping.

The modulation encoding signal could possibly be accounted for by a single auditory nerve fiber with a specific characteristic frequency in the cochlea. For example, the firing of an auditory nerve with a characteristic frequency of 440 Hz will be modulated by both stimulus amplitude transients (f_{AM}) and frequency transients (f_{FM}), resulting in a co-modulated signal (we could also call it a 'beating' signal) that contains a two-sideband pattern in its spectrum, as we observed in this experiment. However, this interpretation cannot account for many other aspects of our results here. First, the resultant modulated response would be a purely amplitude-modulated signal, while we observed pure phase modulation for low f_{FM} stimulus conditions. Secondly, this account cannot explain the observed coding transitions.

Note also that the channels selected per subject for further analysis are based on a single criterion, the amplitude of aSSR at f_{AM} of 37 Hz, and not on any specific sideband frequencies. In addition, the selected channels show a bilateral pair of

origins in auditory cortex, confirming that the results reflect the activities there (Figure 2b).

2.4.3 Neural modulation encoding

The two most simple modulation encoding types are amplitude modulation (AM) encoding and phase modulation (PM) encoding: these arise naturally when the aSSR amplitude or phase depend on the carrier frequency of the stimulus, and so would be expected to occur when the carrier frequency is modulated. Correspondingly, neurons employing AM encoding (Luo et al., 2006) and PM encoding (Patel & Balaban, 2004) have both been proposed. Both types of neurons fire spikes that are phase-locked to the stimulus AM (at frequency f_{AM}), but they differ in the way they encode FM features. Specifically, in each f_{AM} cycle, the AM encoding neuron changes its firing rate to represent the carrier frequency, whereas the PM encoding neuron changes its firing pattern in time to represent the carrier frequency. In other words, both the AM and PM neurons employ temporal coding to track the AM feature, but use rate and temporal coding, respectively, to simultaneously represent the FM feature (f_{FM}). These spiking patterns were illustrated in our previous paper (Luo et al., 2006). Gymnarchus, an African wave-type electric fish, provides a natural example of the modulation-encoding neuron we propose. This species needs to compares timing of sensory feedback from its high-frequency electric organ discharges received at different parts of its body surfaces in order to execute 'jamming avoidance'. This timing comparison mechanism is realized by integrating amplitude and phase difference information in the received electrosensory signals.

Interestingly, amplitude-sensitive and differential phase-sensitive neurons were found to project to an overlapping area where neurons respond to simultaneous amplitude and phase modulations (Kawasaki & Guo, 1998).

2.4.4 Coding transitions

We observe a PM-to-SSB transition as f_{FM} increases from 0.3 Hz to 30 Hz. Specifically, stimuli with slow f_{FM} ($\lesssim$ 5Hz) elicit both significantly stronger peaks and robust phase at both upper and lower sideband frequencies, and the encoding-type parameter α is robustly within the PM encoding region ($\sim \pi$). As stimulus f_{FM} increases (5 Hz $\lesssim$ f_{FM} $\lesssim$ 20Hz), only upper sidebands are elicited and have robust phase, whereas the lower sideband decreases and has noisy phase. Correspondingly, the encoding-type parameter α, the sum of phase parameters for the upper and lower sidebands, also becomes noisy and unreliable. We propose the engagement of a subsidiary AM encoding in addition to the already present PM encoding, which combine in such a way as to cancel the lower sideband, and thus accounts for the observed PM-to-SSB transition. Specifically, for stimuli with slow f_{FM}, the neurons rely solely on a PM encoding mechanism to track the AM and FM features simultaneously. As stimulus f_{FM} increases, these neurons also begin to employ an AM encoding mechanism, also co-representing AM and FM features. Then, both AM and PM encoding mechanisms are present, adding constructively (for the upper sideband) and destructively (for the lower sideband) to generate a SSB signal, and used for concurrent encoding.

It is apparent that a PM encoding mechanism, as seen for slow f_{FM} stimuli, requires more temporal precision and resolution for co-representation and thus also more resource intensive than does an AM encoding mechanism, since it changes the spike firing time to indicate the carrier frequency. AM encoding, in contrast, requires less temporal precision and resolution and only roughly needs to change firing rate to track FM. Therefore, it is reasonable that only PM encoding is involved when tracking stimuli with slower carrier frequency dynamics, and that for tracking stimuli with faster FM requiring more neural resources, the relative contribution of the coarser and more economic AM encoding mechanism is increasingly engaged. These findings parallel findings in marmoset, using click train stimuli, where there is a temporal-to-rate coding switch as the click trains became faster (Lu et al., 2001). Our results are also consistent with fMRI experiments (Giraud et al., 2000; Harms & Melcher, 2002) that have documented changes in the shape and magnitude of sustained responses to AM and FM stimuli as modulation frequency increases. Psychophysical studies have also proposed FM-to-AM transduction (Saberi & Hafter, 1995) and two-stage detection (Moore & Sek, 1996) for FM sound perception. Although this body of research refers to pure AM or FM detection, the underlying ideas apply straightforwardly to our hypothesis.

2.4.5 Asymmetry was not due to different background signal-to-noise ratio

An alternative explanation for the asymmetry between upper and lower sideband performance that must be ruled out is the different signal-to-noise ratios at those frequencies. For example, the decreased performance at the lower sidebands for faster

stimuli might be due to the stronger background noise at lower frequencies, where the lower sideband frequency (f_{AM} - f_{FM}) for higher f_{FM} is located, which in turn would lead to asymmetric results. To test this explanation, we analyzed the direct FM aSSR (aSSR at f_{FM}, not at a sideband), using the same amplitude matrix analysis as that of the sideband frequencies. Since some of the target f_{FM} frequencies (0.3 Hz–30 Hz) are located in an overlapping frequency region to that for lower sideband frequencies with deteriorated performance, if the poor performance for those lower sidebands was due to noisier background, the background noise should also influence the performance of the aSSR at f_{FM}. However, we did not find any decreased performance at this region; on the contrary, the responses at those f_{FM} frequencies were strongly elicited. Therefore, the very same analysis on the same frequency region but using different criteria leads to diverse results, suggesting that it was an encoding transition rather than changes in signal-to-noise that accounted for the asymmetry between upper and lower sideband performances.

2.4.6 Neurons performing specific phase delay

In our neuronal model, AM encoding signals required a specific phase relationship (90-degree phase shift) with PM encoding signals, which accounts well for the observed SSB signals. Neurons with a specific phase shift relative to other neurons have been observed in several studies. For example, in a sound localization study by Fitzpatrick et al. (2000), three types of ITD-sensitive neurons in the inferior colliculus (IC) were found (peak-type, trough-type, intermediate-type); these differ in their characteristic phase, even when they have the same characteristic delay. In other

words, their response patterns are phase-shifted versions of one another. In an ITD discrimination model study, Hancock and Delgutte (2004) have also proposed that the involvement of a phase shift mechanism in a system of solely internal delays could predict psychophysical performances more accurately. In the visual domain, 'lagged cells' have been reported in both LGN and V1 (Saul & Humphrey, 1990; DeValois et al., 2000; Saul et al., 2005). These cells show a specific lagged phase (e.g. by 90°) in their responses compared to ordinary 'non-lagged cells' and are argued to solve the problem of encoding long and variable delays since a given phase difference provides longer time differences at low frequencies. Thus, the hypothesized phase shift in our model is not unrealistic, and most importantly, constructing an additional phase-shifted version of the encoding signal using a different coding scheme (here AM modulation encoding) seems to be an efficient way to establish another dimension of representations of periodic modulation features.

2.4.7 Relationship with systems neuroscience

Brain rhythms are widely studied and are argued to have important functions in the cerebral cortex (see review by Sejnowski & Paulsen, 2006). It has been suggested that gamma-band oscillations (~ 40 Hz) may solve the binding problem (Llinas & Ribary, 1993; Singer & Gray, 1995; Bertrand & Tallon-Baudry, 2000) by synchronously referring diverse fragmented sensory feature representations into a coherent temporal framework to achieve a single cognitive state. This 40 Hz oscillatory activity has been proposed to result from the resonant properties of the thalamocortical system (Llinas, 2000) and interactions between excitatory and inhibitory neurons (Freeman, 2000). It

has also been suggested that the elicited aSSRs reflect the resetting of this 40-Hz brain rhythm by transients in the sensory input, which could also explain the maximum aSSR for modulation frequency around 40 Hz across various stimulus types (Stapells et al., 1984; Rees et al., 1986; Picton et al., 1987; Regan, 1989; Ross et al., 2000). A study of the aSSR to pure AM sound (Ross et al., 2000) systematically examined the effects of stimulus properties (modulation frequency, carrier frequency) on the aSSR (amplitude and phase). Their results suggest that properties of the 40-Hz brain oscillation are modulated by the incoming sensory stimulus. Therefore, an alternative explanation for the encoding transition from the perspective of systems neuroscience is that it reflects a transition between brain states. Specifically, for stimuli with slow dynamics, the brain's 40 Hz rhythms are set by changes in the amplitude of the stimulus envelope (at rate f_{AM}), and the timing of this resetting response, which is reflected in the starting phase of this signal, depends on the fine structure of the incoming stimulus, here, the carrier frequency. As the stimulus fine structure modulations become faster, there arises a more complex pattern in the brain oscillations, which are still reset by envelope changes, while both the resetting gain and resetting phase depend on the incoming stimulus carrier frequency.

2.5 Summary

In this set of studies, we probed the human auditory system with stimuli specifically designed to have two important properties shared by natural communication sounds: First, they are temporally modulated. As introduced

previously, temporal modulations are important physical aspects of communication sounds, and are capable of characterizing sound dynamics. They are found in a wide range of species-specific vocalizations in both animals and humans, and are well represented and preferentially responded to by auditory cortical neurons. They were also shown to be critical for the intelligibility of human speech in speech recognition studies. Secondly, unlike most auditory studies, which traditionally use only AM *or* FM stimuli, our stimuli have simultaneous sinusoidal AM and FM. This co-modulation characteristic is also consistent with properties of natural stimuli in that most natural communication sounds (e.g., human speech, marmoset calls, bird songs, etc.) contain simultaneous temporal modulations in both amplitude and frequency; in other words, AM and FM always co-occur and are inseparable acoustic features of an auditory object. Therefore, in order to ensure that our results are more directly applicable to the auditory system, we designed and employed dynamic stimuli with the simplest forms of these crucial properties of natural communication sounds.

The fundamental motivation underlying these studies is that since AM and FM are inseparable and simultaneous acoustic features of an auditory object, they should be co-represented to achieve 'perceptual unity' of the incoming sound, and we were seeking such a 'co-representation' or 'binding' mechanism to perceptually unify AM and FM acoustic features. By performing a series of data analysis procedures in the spectral domain, we confirmed that *'modulation encoding'* is the fundamental mechanism to co-track and co-represent the two temporal modulation features simultaneously and unify them.

Furthermore, we examined the possibility of a coding transition at higher modulation rates, motivated by temporal-to-rate coding transition findings in neurophysiological studies. Interestingly, we confirmed a smooth transition in the recorded MEG response, from pure phase modulation encoding to a single-upper-sideband-only response pattern (SSB) in the response spectrum.

What is the corresponding neural mechanism underlying this observed PM-to-SSB transition? From a pure engineering point of view, as shown in the results of our simulation, the additional involvement of amplitude modulation encoding responses to ongoing phase modulation could account for this transition. Still, what does that mean? Is there any more explicit explanation for such a transition? What do the underlying neuron populations do to encode these types of dynamic stimuli? Do these MEG results provide any predictions for neurophysiological studies?

This series of questions is deeply related to the various perspectives on MEG and the relationship between macroscopic and microscopic activities, as was discussed largely in chapter 1. The first view, a more intuitive and simple one, is that the modulation-encoding signals recorded with MEG reflect the spiking activity of underlying neuron groups in a relatively linear and direct way. For example, on this view, the interpretation of an aSSR at the corresponding modulation frequency in response to a pure AM sound is that a neuron fires spikes that are phase-locked to the amplitude transients. However, many single-cell recordings in cortical neurons could support this explanation only for lower modulation rates, despite MEG/EEG studies that showed aSSRs at much higher modulation frequencies, up to 100 Hz. Another interpretation is that although individual neurons cannot track fast AM sounds, the

neural population may achieve this by combining the efforts of many neurons. This is still a linear interpretation of the aSSR; it interprets the MEG/EEG signal as the linear sum of the output of a group of neurons, each of which fires spikes that partly track transients. In addition to this limitation, many aSSR studies using different stimuli and in different sensory domains found a common phenomenon that the maximum aSSR was elicited when the modulation frequency was around 40 Hz. This finding cannot be interpreted using the 'linear' view. The hypothesis that posits separate AM and PM neuron groups (see Figure 2-1) also belongs to the linear view in that the 'modulation encoding' refers to the encoding mechanism of a single neuron or groups of neurons, and their spiking activity results in the observed modulated signals in MEG responses.

The second view is more or less from an engineering perspective and regards the brain as a passive signal processing box with a specific impulse response. Each rising edge of the stimulus signal envelope will trigger an intrinsic middle latency waveform pattern with frequencies around 40 Hz, and the aSSR is generated by periodic superposition of those transient responses, as first hypothesized by Galambos (1981). This explanation could account for the aSSR amplitude peak at 40 Hz, but it is not compatible with several experimental results. As stated in Ross et al. (2005*)*: *"For example, superimposed evoked responses could not explain either the time course of ASSR onset or the frequency dependencies of ASSR. The long-lasting ASSR perturbation that was induced by the omission of a single click, in a series of 40-Hz click stimuli, also could not be explained by superimposition of the response to a single click"*.

The final view is different from the first view in that it seeks to explain the MEG/EEG signal as an 'indirect' reflection of the activity of underlying neural ensembles rather than as a simple sum of neural spikes. Instead, on this view, these recorded macroscopic activities reflect dynamic interactions within and across neuron groups and have a very distinct format. This view is also distinguished from the second view in suggesting that the aSSR reflects *induced* activity rather than the superposition of repetitive *evoked* responses, and is facilitated by rhythmic stimulation at frequencies close to the best responding frequency of the underlying neural network (Ross et al., 2005). The fundamental building blocks of these system responses are continuous oscillations at various frequencies, which are modulated according to incoming stimuli. Correspondingly, the observed aSSR response at particular stimulus modulation frequencies actually reflects these stimulus-driven oscillatory brain activities rather than evoked responses (Ross et al., 2005). Therefore, from the perspective of systems neuroscience, an alternative explanation for the observed encoding transition is that it reflects a transition between brain states. Specifically, for slow modulated stimuli, the 40 Hz brain rhythms are partially phase reset by stimulus amplitude transients, and the timing of this resetting response depends on the carrier frequency of the incoming stimulus. For fast modulated stimuli, the reset response shows a more complex pattern dependent on the carrier frequency (fine structure) and modulated both in terms of gain and time.

Chapter 3: Tracking natural speech sentences

3.1 Introduction

How natural speech is represented in the auditory cortex is the most crucial question in cognitive science. We are in a world full of natural sounds, the most important and frequent one among which is natural human speech. From evolutionary perspectives, the brain mechanisms underlying auditory sound processing should evolve and be optimized to process natural speech signals efficiently. Therefore, understanding the neural mechanism of speech processing in the human brain would yield valuable knowledge about the principles of the human auditory system. In addition, in the speech processing field, more efforts have been made to treat speech signals as pure engineering signals composed of many acoustic features and to seek the most behaviorally relevant feature through extensive mathematical computations. The main application of the findings in this field is in telecommunication, where speech signals are manipulated, stored, transmitted and recovered using the smallest possible coefficients while maintaining intelligibility and high resistance to background noise. Understanding the real biological mechanism of speech processing in the human brain could provide direct information about the way the brain encodes speech , leading to great advances in many practical fields, such as telecommunication, artificial intelligence, and others.

However, understanding how natural speech is represented in the auditory cortex also constitutes a major challenge in auditory neuroscience. Human speech is a complex auditory signals that require specification in many dimensions in order to be fully described and represented, and so the crucial question becomes, which dimension is the one the brain utilizes and counts on to process speech? As we introduced previously in detail, speech signals are not stationary—they change over time in both amplitude and frequency. These temporal modulation features have been shown to be closely related to speech intelligibility in speech recognition studies (Shannon et al., 1995; Smith et al., 2002; Zeng et al., 2005) and may be the main dimension in terms of which speech signals are represented in the brain. However, it is still very difficult to explore this question, because the dynamics of natural speech signals are complex, non-regularized, and unpredictable to some extent. Most importantly, we have not yet identified the dimension along which brain activities (specifically those recorded via neuroimaging techniques) should be examined to sort out the neural correlates of speech processing, especially the processing of ongoing speech.

In response to these challenges, the 'reductionist view' has been widely used to try to solve the problem gradually. By 'reductionist view', I refer to probing the auditory system with simplified versions of natural speech that contain one or more regularized and manipulated fundamental acoustic properties. For example, commonly used stimuli are pure tones, FM sweeps, amplitude-modulated sounds, frequency-modulated sounds, etc. Reductionism is closely tied to the field of speech recognition, insofar as studies in the auditory neuroscience domain mainly focus on

those acoustic features that are found to be crucial to speech recognition. The 'reductionist view' is also the main perspective most neuroimaging studies adopt to find the neural correlates of speech processing, both spatially and temporally. In PET/fMRI studies, seeking the 'speech-specialized' areas on hemodynamic spatial maps is the main focus, and finding such an area is tantamount to a conclusion. Most of these studies compared speech with non-speech, and although this paradigm could possibly answer the question of *where* these speech signals are processed, the results still lack very crucial information due to the limitation of this technique: *How* are these speech signals processed? *How* are different speech signals represented differently? In other words, a complex question has been reduced to a 'where', the answer to which will be the same for all speech; furthermore, the processing mechanism for dynamic properties of speech has been neglected. MEG and EEG, as neuroimaging techniques with high temporal resolution, are ideal tools for exploring this question, but so far too much attention has been paid to localizing a dipole at a single time point.

This complex question has also been investigated intensively in animal neurophysiology using species-specific communication sounds (Nelken et al., 1999; Wang et al., 2003; Nelken 2004; Machens et al., 2003, 2005; Woolley et al., 2005; Narayan et al., 2006). The advantage of single-cell recording is its good resolution both spatially and temporally. Spatially, some neurons are found to be selective to species-specific vocalizations and not others. For example, it has been shown that natural vocalizations of marmoset monkeys produce stronger responses in A1 than do spectrally similar but temporally altered vocalizations (Wang et al., 2003). When

testing responses to these sound pairs in the auditory cortex of the cat, whose A1 shares similar basic physiological properties with marmosets, neurons in the cat A1 did not differentiate the natural marmoset vocalizations from their time-reversed versions. These observations suggest that this constructed selectivity of cortical neurons depends on the behavioral relevance of these signals in the species' environment and on learning-induced cortical organization. However, although regions found to selectively process conspecific call sounds have been confirmed in many animals, it is not known precisely 'what' (stimulus properties, recognition of acoustic objects or call-sound meaning) is represented and 'how' different conspecific calls are discriminated in the responses. Temporally, the neuron spiking pattern in response to these communication calls is examined and compared to the spiking pattern to call-like or non-call-like complex sounds. Many studies showed that single auditory neurons or neuron groups fire relatively temporally precise and similar spike patterns across trials, compared to other formats of complex sounds. In an experiment by Hsu et al. (2004) investigating neural encoding of songs by single neurons in zebra finches, they calculated the mutual information contained in the time-varying mean firing rate of the neural responses, and compared song, song-like synthetic sound, and non-song-like synthetic sound. They found that the songbird auditory system showed selectivity for song and song-like sound, and the corresponding spike trains carried more information than other formats of complex synthesized songs in that the temporal spike trains were more similarly precise across trials. Machens et al. (2003) investigated whether single auditory neurons in Chorthippus biguttulus could discriminate conspecific communication signals, by performing a classification

analysis based on the temporal spike train response to different communication signals. They confirmed that information sufficient to distinguish songs is readily available at the single-cell level when the spike trains are analyzed on a millisecond time scale. A similar analysis was performed on song birds to investigate the song discrimination ability of single cortical neurons, and this ability was confirmed again when spike trains were analyzed on the millisecond temporal scale (Narayan et al., 2006). Together, these observations suggest that species-specific communication sounds are well represented and processed in animal auditory cortex by some neurons that act as 'call-detectors', and these neurons fire temporally precise and stable spike patterns to their preferred conspecific sounds. In other words, specialized neurons are representing and encoding communication sounds, and the information that allows discrimination of different communication sounds is carried in their temporal spike patterns.

Many brain imaging studies have showed that some areas of the brain are significantly associated with speech processing, and the elicited cortical brain signals provided valuable information about the spectral and temporal processing of speech stimuli (Suppes et al., 1999; Giraud et al., 2000; Ahissar et al., 2001; Zatorre et al., 2002; Griffiths et al., 2004; Boemio et al., 2005; Luo et al., 2005, 2006; Scott et al., 2006). However, what exactly are the components of these macroscopic brain activities that can reliably track and discriminate speech sentences? This question remains unanswered, due to a dearth of knowledge about the relationship between macroscopic and microscopic activities.

We hypothesized that the phase pattern of cortical rhythms might be a possible encoding mechanism to track ongoing natural speech signals. Our hypothesis was motivated by several findings. First, as observed in single-cell studies, cortical neurons or neuron groups fire different spike patterns in response to different communication sounds, thus demonstrating that temporal information is crucial to this discrimination ability (Machens et al., 2003; Narayan et al., 2006). Secondly, failure to find a significant differentiator in the evoked temporal waveform in previous EEG/MEG experiments makes us re-examine ongoing electromagnetic responses using other analysis methods. Thirdly, EEG/MEG signals were found to be mainly dominated by stimulus-induced changes in endogenous brain dynamics rather than by stimulus-evoked brain events, and these inherent brain rhythms have also been found to have functional significance in auditory object perception, and this constitutes a very different activity format compared to results from single-cell studies (Hari et al., 1997; Engel et al., 2001; Makeig et al., 2002; Penny et al., 2002). Therefore, brain oscillations at certain frequencies may be more likely to carry information than the evoked temporal waveform, which comes from the sum of the components of all frequencies. Finally, the phase of ongoing brain rhythms was found to be crucial for tracking ongoing stimulus feature changes; in other words, stimulus-dependent phase modulation of certain ongoing brain oscillations is a very possible mechanism of representation reflected in macroscopic activities. For example, in a MEG study investigating tone sequence tracking in human auditory cortex (Patel & Balaban, 2000), researchers found that the phase of the elicited aSSR at the envelope modulation frequency could reliably track the tone sequence even in single trials, and

that this tracking ability is best for stimuli with the statistical structure of music. In addition, as described in Chapter 2, we confirmed that phase modulation encoding is employed to track fine structure changes in the stimulus. In sum, the ongoing phase pattern of certain frequencies in MEG responses is hypothesized to represent and track speech signal dynamics.

Due to low signal-to-noise ratio in EEG/MEG responses and certain characteristics of these macroscopic activities, it is difficult to describe and test the unique pattern shown in each single trial response, as is possible with neurophysiological experiments. A common way to address this problem is to remove noise bv averaging responses across trials; however, this technique is not applicable here. Instead, I took a simple first step to test this hypothesis, motivated by results from single-cell studies that show that for specific communication sounds, cortical neurons will fire reliable spike patterns in single trials, and different communication sounds will elicit different spike patterns. Correspondingly, if the phase pattern of the MEG response to one natural speech signal in a single trial contains information specific to this speech signal and distinct from that of other speech signals, the cross-trial phase coherence, characterizing how similar the phase patterns of responses are across trials, should be larger for responses to the same speech signal than for composites made up of responses to different speech signals. In addition, motivated by neurophysiology studies that show that communication sounds are processed mainly in auditory cortex, human studies should also find that tracking more or less takes place in the auditory cortex.

Other relevant and important questions: Even if we confirm such a tracking pattern, are these mechanisms specific to processing speech, or are they general mechanisms for representing all auditory sounds? Do these neural correlates relate to behavioral perception, for example, to speech intelligibility (Narain et al., 2006; Scott et al., 2006)?

3.2 Materials and Methods

3.2.1 Subjects and MEG data acquisition

Six native English speakers with normal hearing and no neurological disorders provided informed consent before participating in the experiment. Neuromagnetic signals were recorded continuously with a 157 channel whole-head MEG system (5 cm baseline axial gradiometer SQUID-based sensors, KIT, Kanazawa, Japan) in a magnetically shielded room, using a sampling rate of 1000 Hz and an online 100 Hz analog low-pass filter, with no high-pass filtering.

3.2.2 Stimuli

Three natural speech sentences (“It made no difference that most evidence points to an opposite conclusion."; "He held his arms close to his sides and made himself as small as possible."; "The triumphant warrior exhibited naive heroism.") with sampling frequency of 16 kHz were selected from the TIMIT speech database and their durations were in the range of 4000 ms to 4700ms. For each speech sentence, we constructed 4 types of speech-noise chimaeric stimuli (Env4, Fin1, Env1, Fin8), the

spectrogram of which were shown in Figure 3-1. These speech-noise chimaeras contain speech information in either their envelope (ENV) or their fine structure (FIN) and another important variable is the number of frequency bands used to split sound (See Smith et al., 2002). The intelligibility scores for Env4, Fin1, Env1 and Fin8 were shown in a previous behavioral study to be 0.85, 0.7, 0.05 and 0.2, respectively (Smith et al., 2002). Correspondingly, they can be separated into 'intelligible speech stimuli' containing original, Env4 and Fin1, and 'unintelligible speech stimuli' containing Env1 and Fin8. The whole stimuli set were then amplitude modulated at 50 Hz.

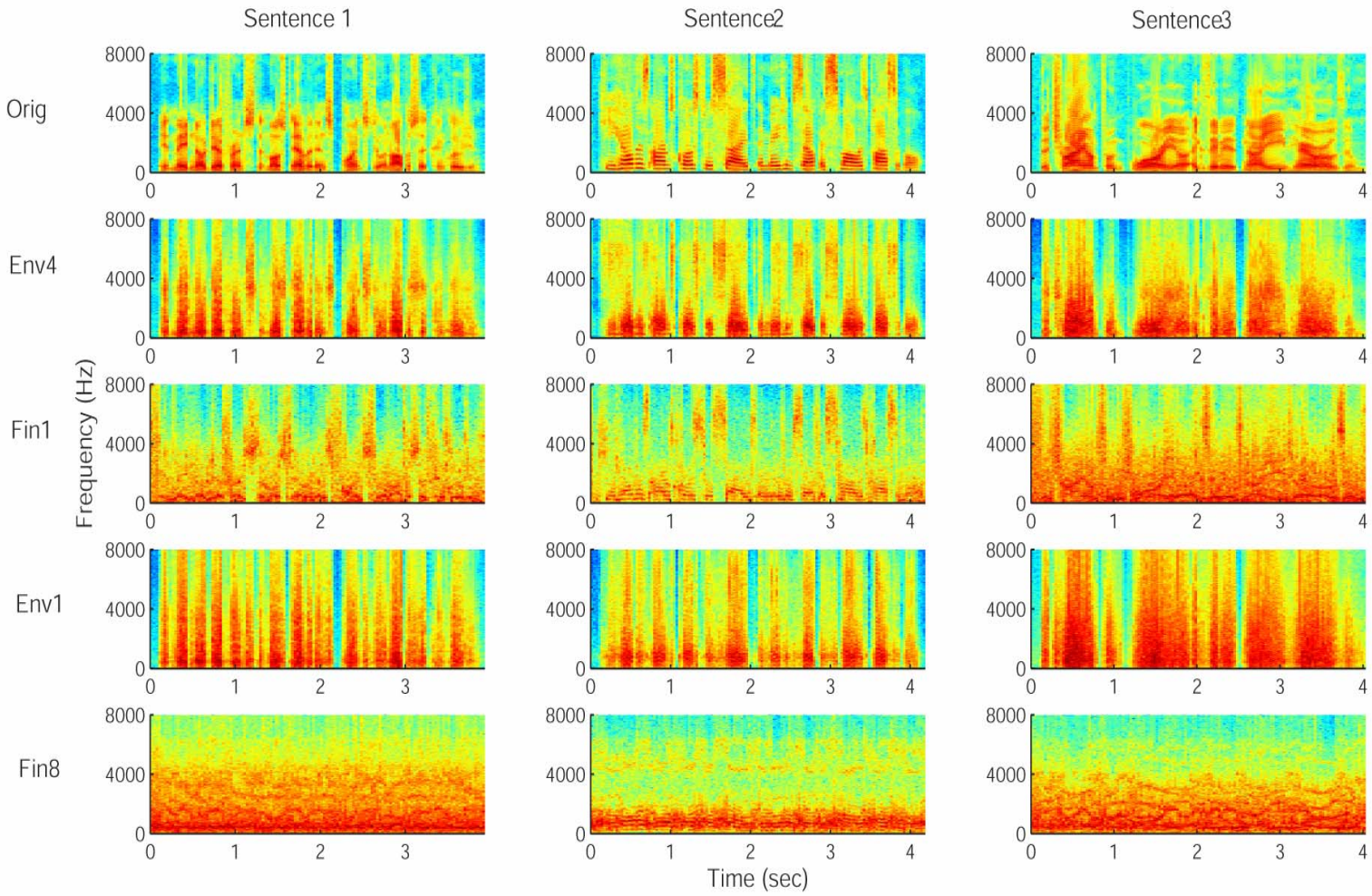

Sentence 1
Sentence2
Sentence3
Orig
Env4
Fin1
Env1
Fin8
Frequency (Hz)
Time (sec)
8000
4000
0
0
1
2
3
4

Figure 3-1. Spectrograms of sentence materials and their manipulated versions. Each of the three original sentences was turned into chimaeral stimuli. *Orig*, *Env4*, and *Fin1* are intelligible at different levels (100%, 85%, 70%, respectively); *Env1* and *Fin8* are not intelligible (and were always used as the second sentence in a given trial). Only intelligible sentences were analyzed.

3.3.3 Experiment procedures

In an initial pretest, the participants were presented with 1 kHz tone pips (duration 50 ms) to determine their M100 evoked responses. Subjects were then told to listen to the (original and degraded) versions of spoken sentences. On each speech trial, two sentences were presented sequentially with 1-sec interval between them; subjects were instructed to indicate by button-press whether they were same or different sentences. The first one was always drawn from the intelligible set (original, Env4, Fin1), the second one was always unintelligible (Env1, Fin8). Each of the nine intelligible conditions (3 sentences, 3 intelligible conditions) was presented 21 times at a comfortable loudness level (~70 dB). Eleven other duration-matched sentences from the TIMIT database were selected and their unintelligible versions (Env1, Fin8) were constructed. These unintelligible speech stimuli were randomly selected as the second stimulus in each speech trial. Only cortical responses to intelligible stimuli were extracted for further analysis.

3.3.4 Data analysis

'Across-group' signal construction

All response trials (21 here) for the same original speech stimulus are termed as 'within-group' signals (3 'within-group' signals corresponding to 3 original sentences). Then, 7 response trials (one third of the total 21 trials for each stimulus condition) are randomly chosen from each of the 3 'within-group' signals and combined to construct a 21-trial 'across-group' signal. Three 'across-group' signals are constructed by repeating the random combination procedure 3 times.

Dissimilarity function

For each of the six 21-trial signals (3 'within-group' and 3 'across-group' signals), the spectrogram of the first 4000 ms of each single trial response was calculated using a 500 ms time window in steps of 100 ms for each of the 157 MEG recording channels, and the calculated phase and power as a function of frequency and time were stored for further analysis. The 'cross-trial phase coherence ($Cphase$)' and 'cross-trial power coherence ($Cpower$)' were calculated as:

$$Cphase_{ij}=\left(\frac{\sum_{n=1}^{N}\cos(\theta_{nij})}{N}\right)^2+\left(\frac{\sum_{n=1}^{N}\sin(\theta_{nij})}{N}\right)^2$$

$$Cpower_{ij}=\frac{\sqrt{\frac{\sum_{n=1}^{N}(A_{nij}^{\ 2}-\overline{A_{ij}^{\ 2}})^2}{N}}}{\overline{A_{ij}}^2}$$

Where θ_{nij} and A_{nij} are the phase and amplitude at the frequency bin i and temporal bin j in trial n, respectively. Both $Cphase$ and $Cpower$ will be in the range

of [-1 1]. Note that larger $Cphase$ value corresponds to strong cross-trial phase coherence, whereas smaller $Cpower$ value corresponds to strong cross-trial power coherence. These calculated cross-trial coherence parameters ($Cphase$, $Cpower$) were compared between each of the 3 'within-group' signals and each of three 'across-group' signals separately. The dissimilarity function for each frequency bin i was defined as:

$$Dissimilarity\_phase_i = \frac{\sum_{j=1}^{J} Cphase_{ij,within}}{J} - \frac{\sum_{j=1}^{J} Cphase_{ij,across}}{J}$$

$$Dissimilarity\_power_i = \frac{\sum_{j=1}^{J} Cpower_{ij,across}}{J} - \frac{\sum_{j=1}^{J} Cpower_{ij,within}}{J}$$

The resulted 3 dissimilarity functions (3 'within-group' - 'across-group' pairs) were averaged, and in results, each of the 157 MEG channels has two dissimilarity functions as a function of frequency ($Dissimilarity\_phase$, $Dissimilarity\_power$), in which the value significantly above 0 indicates larger cross-trial coherence of 'within-group' signal than that of 'across-group' signal.

Phase dissimilarity distribution map

The $Dissimilarity\_phase$ function was then divided into 5 canonical electrophysiological frequency bands (Theta: 4~8 Hz; Alpha: 8~14 Hz; Beta1: 14~20 Hz; Beta2: 20~30 Hz; Gamma: 30~50 Hz) and the average values within each frequency band was calculated, resulted in 5 $Dissimilarity\_phase$ values for the 5

frequency bands respectively. 'Phase dissimilarity distribution map' for the 5 frequency bands were then constructed separately in terms of the corresponding *Dissimilarity _ phase* value of all 157 channels in this frequency band, and drawn as a spatial map with large values represented by stronger red color and small values represented by stronger green color. For comparisons, for each subject, the pretest M100 responses at the latency of M100 were also extracted and the absolute values of all 157 MEG channels were drawn as a spatial map.

Channels selection

For each subject, 20 channels with maximum *Dissimilarity _ phase* value in Theta band (4~8 Hz) were selected for further classification and grand average analysis.

Classification performance

The classification analysis was performed on the selected 20 channels with maximum theta phase dissimilarity values, for each of the 6 subjects separately, to verify whether the theta band phase pattern is sufficiently robust to discriminate among the sentence stimuli in single MEG response trials. For each sentence, the 'theta phase pattern' as a function of time for one single trial response under one sentence condition was arbitrarily chosen as a template response for that sentence. The 'theta phase pattern' of the remaining trials of all conditions was calculated and their similarity to each of the 3 templates was defined as the distance to the templates. The single trial response was then classified to the closest sentence template. The

classification was computed 1000 times for all the 21 trials in each stimulus condition, and for all the selected 20 channels in each subject, by randomly choosing template combinations. The classification results were then averaged to be in the range from 0 to 1, indicating the percent that an empirical single-trial response to a specific stimulus condition is classified to one stimulus condition.

For the 9-condition classification analysis, because of the large computation requirement, the classification was only computed 200 times by randomly choosing template combinations.

3.3 Results

3.3.1 Theta-band phase pattern could discriminate speech signals

To investigate whether information in the electrophysiological responses can be relied on to discriminate different speech sentence stimuli, we developed an analysis that could easily test the cortical activity patterns relevant to the representation of specific sentences in single trials. We call the response to trials for the same sentence conditions as 'within-group' signals. Correspondingly, we constructed 'across-group' signals by randomly mixing trials from different stimulus conditions (Figure 3-2a). The logic here is that if the phase pattern at specific frequencies successfully discriminates between sentences, as we hypothesized, the phase patterns of 'within-group' signals should be more similar across trials than that of 'across-group' signals. In turn, the cross-trial phase coherence of 'within-group' signals should be larger than that of 'across-group' signals, because the response trials of formal signal contain relatively similar encoding activities for a specific same speech stimulus, whereas the

response trials of the latter signals are responses to different speech stimulus. On the other hand, if the hypothesis is wrong and there are no common activities across response trials to the same sentence, there should be no difference between the 'within-group' and 'across-group' signals. The differences between these two types of signals were characterized by 'Phase dissimilarity function' (Figure 3-2b).

We observed well-defined peaks in the 4-8 Hz frequency range in this 'phase dissimilarity function' in many channels (Figure 3-2b, upper row), indicating that the phase pattern in the theta band (4-8 Hz) rather than other frequencies could discriminate between the different sentence stimuli. The averaged 'phase dissimilarity function' across 20 selected channels representative of tracking ability confirms the role of theta band (Figure 3-2 b, right). To assess whether the observed phase-based discrimination ability is accompanied by corresponding discrimination ability in the power of the theta band response, we calculated the 'power dissimilarity function,' characterizing the difference in the across-trial power coherence between 'within-condition' and 'across-condition' signals. There were no significant peaks in this analysis (Figure 3-2b, bottom row), confirming that stimulus discrimination is based on pure phase information.

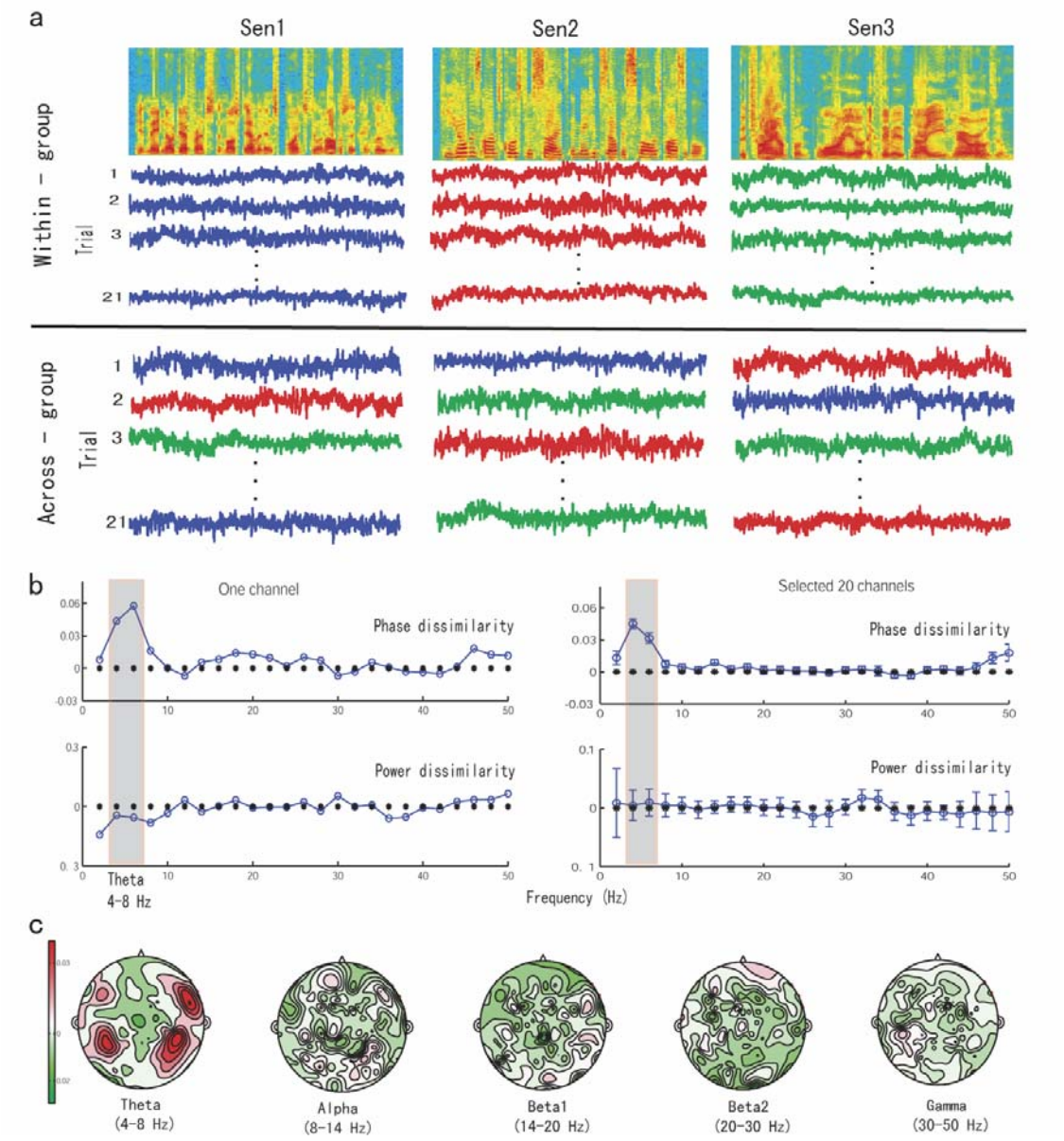


Figure 3-2 Spectrograms of sentence stimuli and representative MEG data for one subject. a, Example stimuli and single-trial responses (blue, red, green) from one channel. 'Within-group' bins (same color) constitute responses to the same condition, 'across-group' bins (mixed colors) to a random selection of trials across conditions. b, Left: 'Phase dissimilarity function' (upper) and 'Power dissimilarity function' (lower) as a function of frequency (0-50 Hz) for the same example channel. Grey box denotes the theta range (~4-8 Hz) where the 'phase dissimilarity function' shows peaks above 0. Right: averaged dissimilarity functions across 20 selected

channels showing maximum phase dissimilarity values in theta band for same subject (mean and standard error). c, 'Phase dissimilarity distribution map' for 5 frequency bands in same subject. Channels depicted with stronger red colors represent large phase dissimilarity values. The 'theta phase dissimilarity distribution map' shows the 'dipolar' distribution typical of auditory cortex responses.

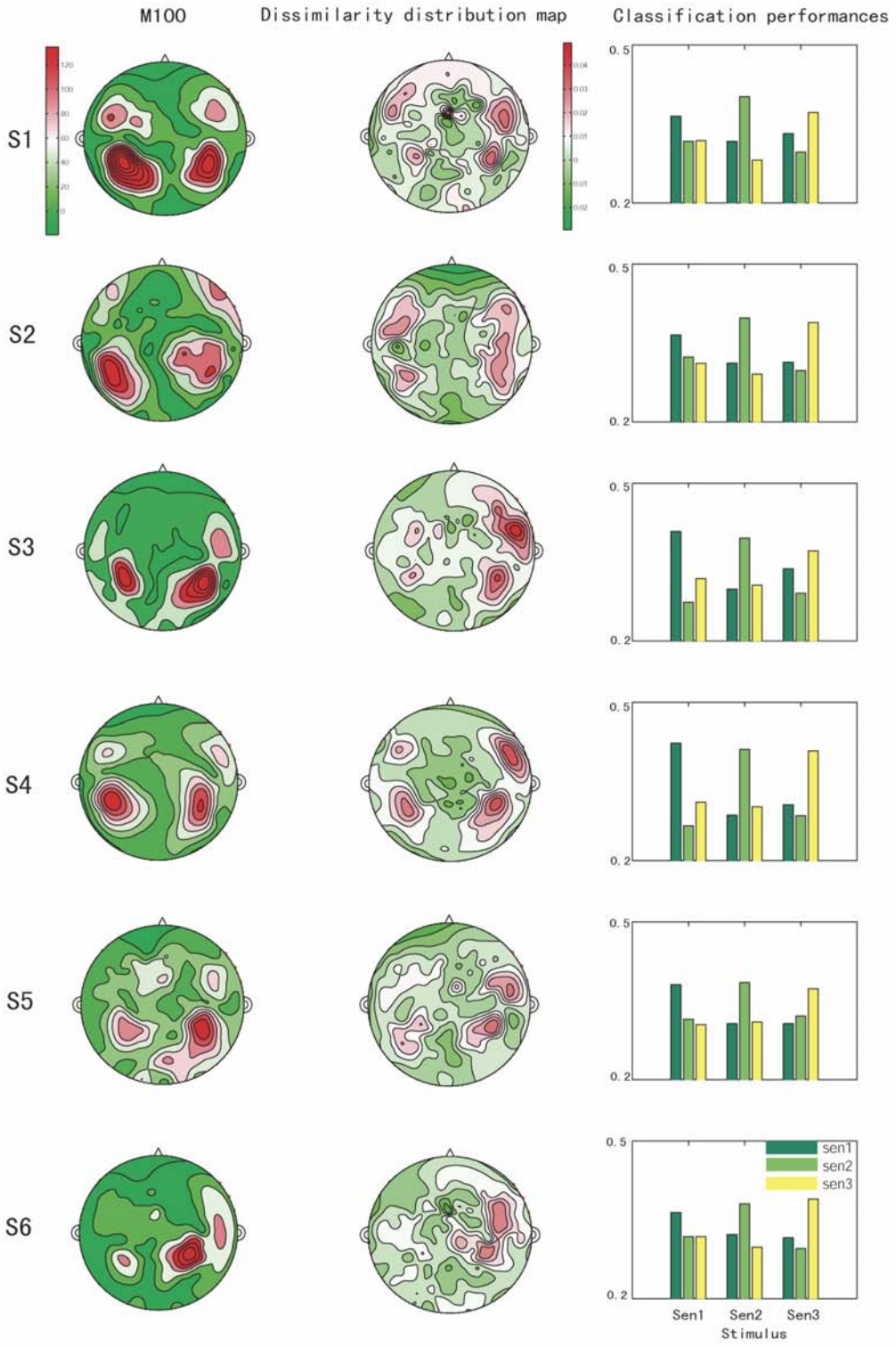
3.3.2 Auditory cortex origin of Theta-band phase tracking

We examined the corresponding spatial distributions of such 'theta phase dissimilarity function', by drawing a spatial map indicating the dissimilarity values of all 157 MEG channels. Interestingly, the 'theta phase dissimilarity distribution map' showed a clear auditory cortex origin (Figure 3-2c). The spatial distributions for other frequency ranges were noisy and not indicative of localized underlying activity (Figure 3-2c). This analysis strengthens the argument that it is the phase of theta band activity in auditory cortex that tracks the sentential stimuli.

Crucially, a 'theta phase dissimilarity distribution map' with auditory origin was observed in every subject (Fig 3-3, middle). For comparison, the contour maps for the M100/N1m, the largest and most robust auditory response originating in superior temporal cortex, are shown for each subject (Figure 3-3, left). This response is generated in superior temporal cortex roughly 100 ms after sound onset [24] and was elicited here in a pretest using 1 kHz pure tones pips. Despite large differences in response amplitude, the two spatial maps show a good spatial match, confirming the auditory cortex origin of the theta-band phase pattern. Note that the 'theta phase dissimilarity distribution map" (Figure 3-3, middle) also shows right hemisphere lateralization. For each subject, the 20 MEG channels with the largest theta phase

dissimilarity were selected for further analysis and are the channels with stronger red color.

Figure 3-3 Auditory cortex identification, 'theta phase dissimilarity distribution map,' and classification performance for all subjects. Left: M100 contour map for each subject. Red indicates large absolute response value at M100 peak latency. Middle: Theta phase dissimilarity distribution map. Right column: Classification performance. The horizontal axis represents the stimulus condition (Sen1, Sen2, Sen3) and the bar color represents the category (Sen1, Sen2, Sen3) this stimulus was classified to. The height of the bar represents the proportion that one single-trial to this stimulus condition (horizontal axis) was classified to this stimulus category (bar color). Note that the sum of the three clustered bars is 1.


M100
Dissimilarity distribution map
Classification performances
S1
S2
S3
S4
S5
S6
0.5
0.2
sen1
sen2
sen3
Sen1
Sen2
Sen3
Stimulus

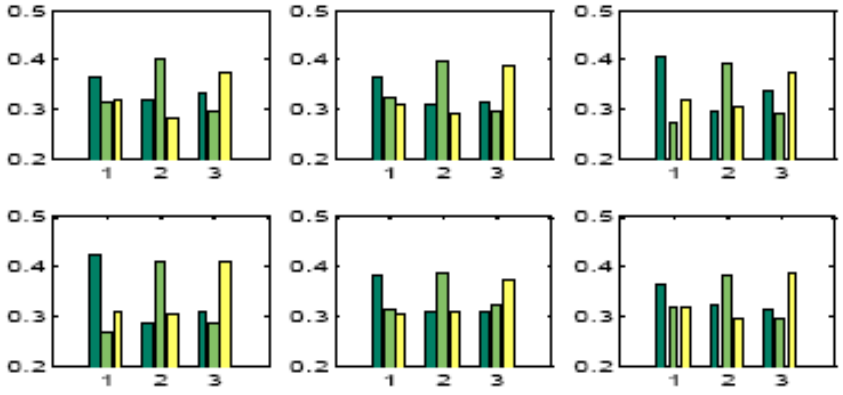
3.3.3 Classification performances

Strikingly, the data from all subjects showed good classification performance (Figure 3-3, right). For each of the 3 sentences, trials were classified with higher proportion into the correct category than not, indicating that the 'theta phase pattern' could be relied on for sentence discrimination in single trial responses.

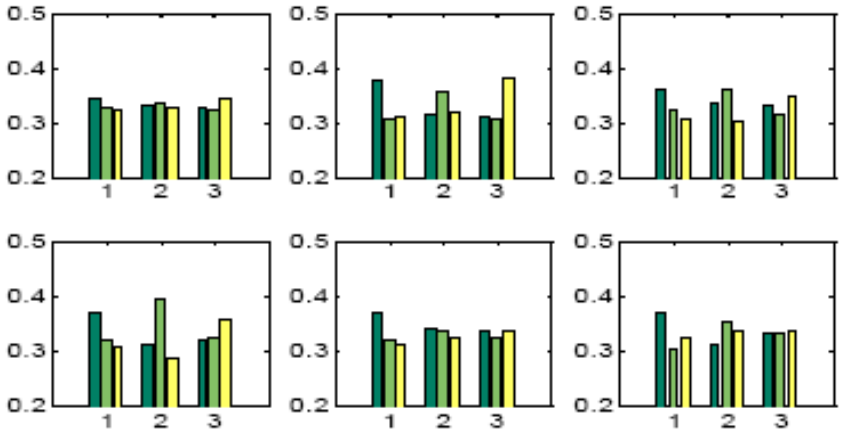
3.3.4 Discrimination ability correlates with speech intelligibility

Beyond successful sentence classification, it can be demonstrated that the phase of the theta band response has compelling perceptual correlates. We show that the discrimination ability of 'theta phase pattern' correlates with intelligibility of the speech materials, by performing the same classification analysis on responses to degraded versions of the same sentences - the speech-noise chimaeras. We constructed two chimaeras for each sentence, 4-band chimaeras containing only acoustic envelope information (Env4), and 1-band chimaeras containing only fine structure information (Fin1). Their intelligibility level (proportion correct) is 0.85 and 0.70, respectively, based on previous studies. This analysis reveals degraded classification performance (Figure 3-4, middle, lower) compared to that of the original sentence stimuli (Figure 3-4, upper), although the classification performance for these degraded speeches is still very good in all subjects. The less intelligible a sentence is, the less reliable is the theta phase pattern. Figure 3-5 showed the grand average of classification performance for the three speech versions across the six subjects.

Orig for all 6 subjects:



Env4 for all subjects:



Fin1 for all subjects:

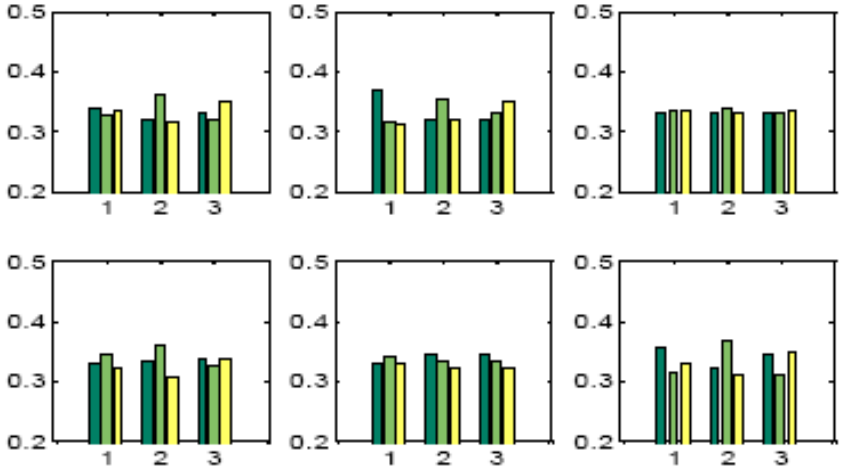


Figure 3-4 Classification performances for all 6 subjects. Upper: Original speech sentence classification. Middle: Env4 speech-noise chimera classification. Lower: Fin1 speech-noise chimera classification

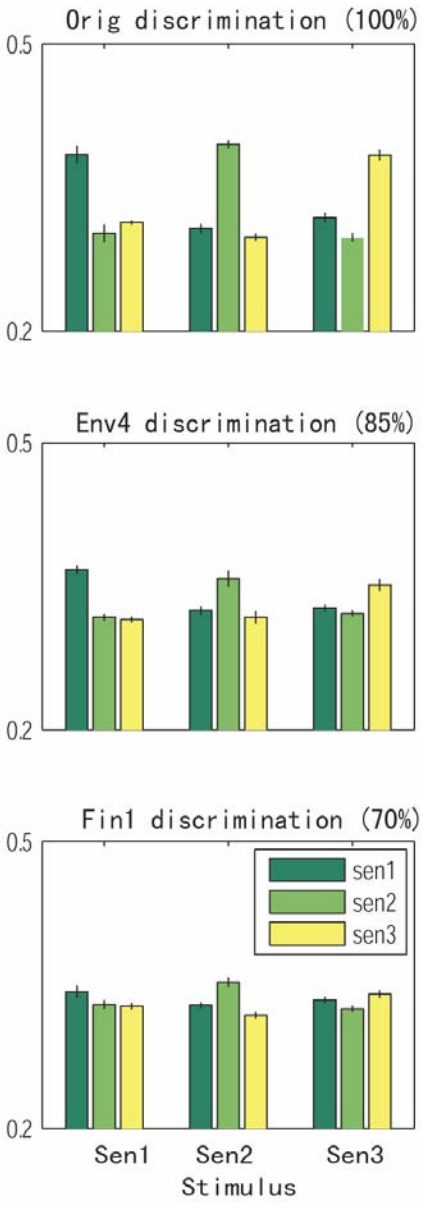


Figure 3-5 Classification performance as a function of intelligibility. Less intelligible stimuli show parametrically degrading classification. Top: Discrimination of 3 original sentences. Middle: Discrimination of three Env4 sentences. Bottom: Discrimination of three Fin1 sentences. The percent value in each figure indicates the intelligibility score from a previous experiment

3.3.5 Category membership

We tested whether the 'theta phase pattern' could reflect 'category membership' of Env4 and Fin1 responses to the corresponding original (undistorted) speech signal by doing the same classification across all nine stimulus conditions (3 sentences × 3 stimulus manipulations). The grand average of the nine-condition classification performance is summarized in a 9-by-9 classification matrix for illustration purposes (Fig. 3-6a). The elements on main and sub- diagonal axes denoted by red lines indicate the correct classification to the stimulus condition itself and the classification to other versions of the same sentence, respectively. These diagonal axes more or less showed peak values. Such 'clustering' of different versions of the same sentence is shown more explicitly in Figure 3-6b. The three versions (Orig, Env4 and Fin1) of each sentence were mostly classified into the corresponding sentence category (rectangular boxes) rather than into other groups. Moreover, among the three versions of each sentence, Fin1 stimuli showed the lowest classification performance, in accordance with the corresponding lower intelligibility scores.

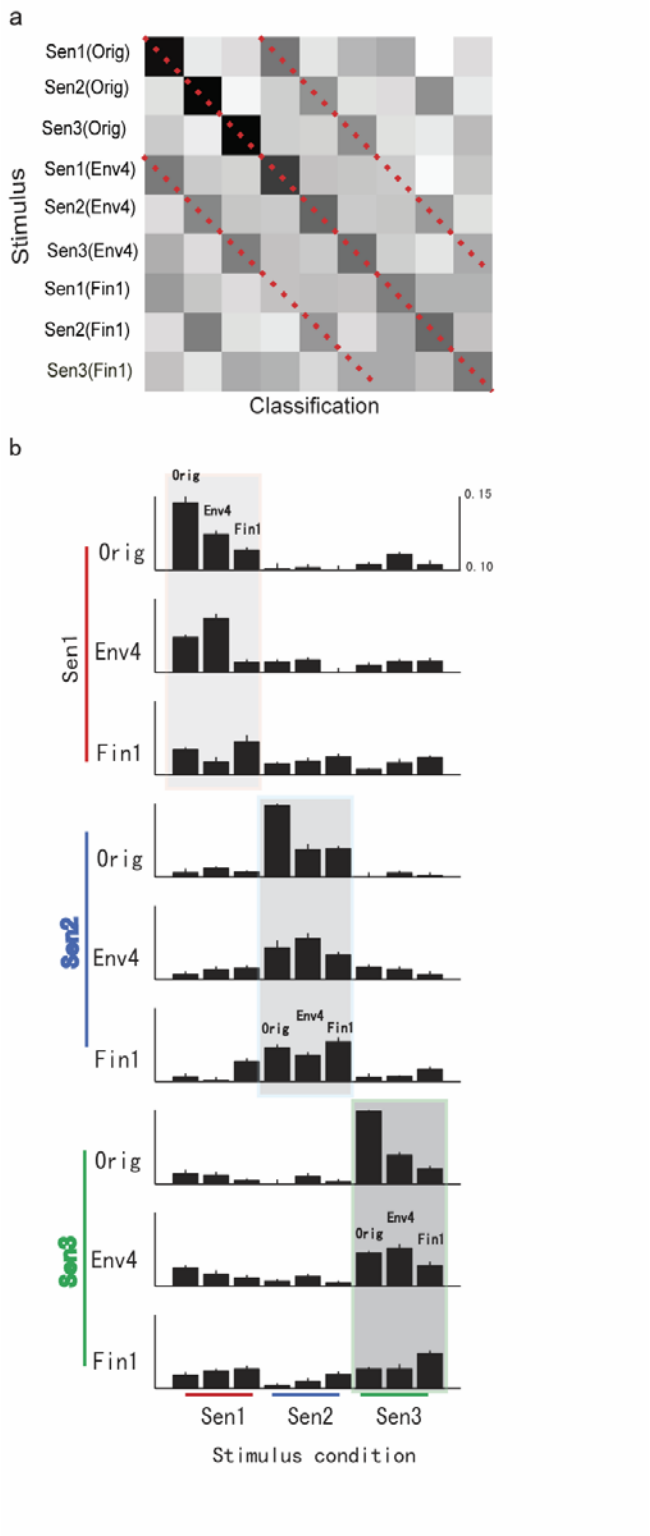


Figure 3-6 'Theta phase pattern' reflects category membership. a, Grand average of 9-condition classification matrix across 6 subjects. Each cell in the matrix represents the percent that a

response trial for this stimulus condition (corresponding row) was classified to this stimulus category (corresponding column). The sum of each row is 1. Red lines indicate the main diagonal and sub-diagonals, where the response was classified to stimulus itself or members in the same category (different versions of same sentence). b, Classification histograms for each of the 9 stimulus conditions (3 sentences × 3 manipulated conditions). Rectangles indicate the range of corresponding correct category membership. For example, for all 3 versions of sentence 1 denoted by red vertical line (upper three rows), the rectangle covers the stimulus conditions all belonging to sentence 1, and should be classified into with higher percent than into other rectangles. Error bars indicate the standard error across 6 subjects.

3.3.6 Classification performance develops over time

We examined the time course of the classification performance in terms of theta band phase pattern in each trial. We extracted the temporal segment (first 500 ms, first 1000 ms, first 2000 ms, first 3000 ms and first 4000 ms) of recorded MEG responses and did the same classification performances as we did before on them separately and compared the classification performances. Interestingly, we confirmed a gradual development of such classification ability based on theta phase pattern. Specifically, the correct classification began to emerge around 2000 ms from the beginning of speech sentence stimulus onset, as shown in Figure 3-7.

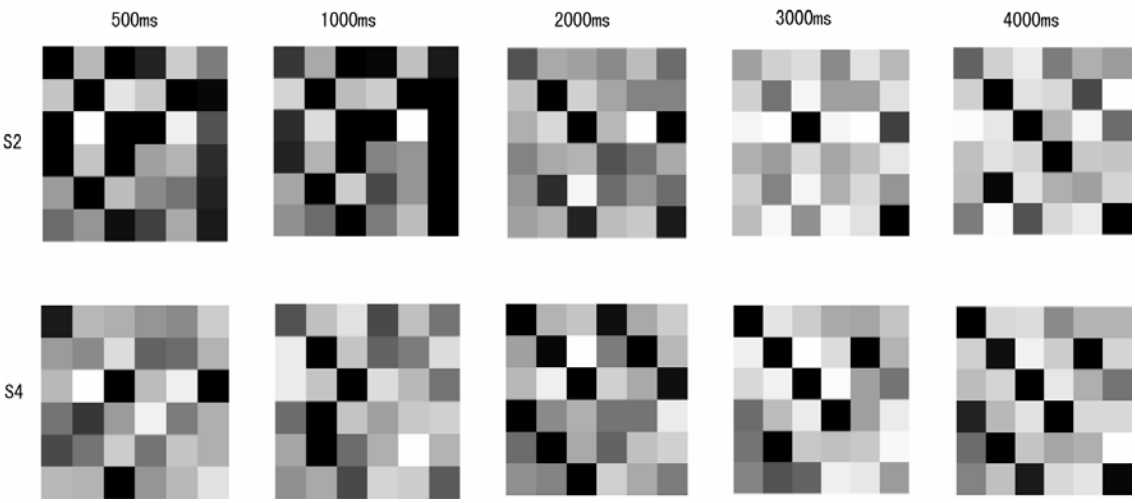


Figure 3-7 Sample classification matrices as a function of integration time for 2 subjects. A six-condition (Original and Env4 versions of 3 sentences) classification analysis is shown. For example, 500-ms classification performance was calculated on only the first 500 ms of response, 1000-ms classification performance was calculated on the first 1000 ms of response, and so on. Unsurprisingly, because of the long period of theta (~200 ms), the MEG-recorded response must be collected over several periods before it becomes a robust discriminator. For Subject 2, robust discrimination ability emerged around 2000 ms, and for subject 4, the discrimination ability emerged around 3000 ms.

3.4 Discussion

3.4.1 MEG data reflect system activities

Different views of macroscopic activities in neuroimaging studies lead to different analysis methods employed to examine MEG/EEG activities. Although in animal neurophysiology studies using species-specific communication sounds, single

neurons were found to fire precise temporal spike patterns for the same stimulus across lots of trials (Mechans et al., 2003; Narayan et al., 2006), such precise and stable response could not be confirmed in MEG/EEG activities. It is often argued that the reasons for these noisy macroscopic responses are the low signal-to-noise ratio of such recording, and the mixture of different temporal spike patterns of underlying single neurons firing. Based on this view, averaging temporal waveform of MEG/EEG response across lots of trials is regarded to be able to more or less remove the noise. In addition, also from the same perspectives, the temporal peaks/troughs in the recorded MEG temporal waveform are regarded as the important information-carrying components at macroscopic level corresponding to those spiking at microscopic level. However, this linear view could not explain many discrepancies between activities at the macroscopic and microscopic level. For example, M100, a MEG big response happened around 100 ms after stimulus onset, is traditionally viewed as onset response and is resulted from the peaky spiking activities of underlying neurons. However, single-cell studies showed that single neurons actually fire spikes only several milliseconds after stimulus onset. Therefore, the temporal peaks/troughs in macroscopic activities (here MEG) could not be simply explained by microscopic single-cell activities, even considering the transfer delay.

On the other hand, in spite of good spatial and temporal resolution in neurophysiological recordings, there still exist many problems that single cell studies could not answer directly or possible, especially at cortical level where neurons seem to do more abstract and complicated work. For example, it is still hotly debated whether there are ‘call specific’ cells in animal auditory cortex, because although

these cells show preference to calls, they seem to be also involved in many other tasks, even in other sensory domains. It has been confirmed that many acoustic features are well represented in some cortical neurons, but the whole picture of the representation of a unified auditory object remains obscure, because of the microscopic-level recording, and it is analogous to 'see trees instead of forest'.

Complementarily, MEG/EEG is an ideal technique to overcome these shortcomings of single-cell studies in animals. We could observe normal human brain activities in real time, non-invasively, and thus could get the most direct information about human brain responses during normal cognitive tasks. Secondly, MEG/EEG activities come from activities and interactions of large numbers of underlying neurons and reflect the system activities or brain states which could not be assessed from single-cell recordings. System activities have been shown to be significantly important in stimulus encoding, category learning, memory in previous studies, and are argued to be more behaviorally relevant. For example, in a neurophysiology studies (Ohl et al., 2001), where Mongolian gerbils were trained to categorization of frequency-modulated sweeps with different sweeping direction ('rising' or 'falling'), electrical activities in the auditory cortex were recorded during this learning process. The behavioral transition to successful learning of categorization could not be observed in activities in single unit. However, when regarding the population activities as a multiple dimension signal space, they showed a strikingly correlation with behavioral transition. In other words, it is the complex and dynamic brain state and population activities that are of close relevance to behavioral perception. Recent advances in neuroimaging have also shown that it is possible to accurately decode a

person's conscious experience based only on employment of pattern-based approaches in non-invasive measurements of their brain activities. For example, multivariate neuroimaing approaches, in contrasts to strictly location-based conventional analyses in fMRI studies, by taking into account the full spatial pattern of brain activities measured simultaneously at many locations, could dramatically increase the information that can be decoded about the current mental states. Recent work demonstrated that pattern-based decoding of BOLD fMRI signals acquired at relatively low spatial resolution can successfully predict the perception of low-level perceptual features, for example, the orientation, direction of motion and even perceived color of a visual stimulus presented to an individual subject (Kamitani & Tong, 2005; Haynes & Rees, 2005). In a fMRI study examining binocular rivalry, by training a pattern classifier to distinguish between distributed fMRI response pattern associated with the dominance of each percept, this classifier could achieve a dynamic prediction of any perceptual fluctuation with high temporal precision (Haynes & Rees, 2005). These fMRI findings also strongly indicate that population activities across multiple locations and dynamic system activities contain important information about an individual's current perception and cognitive state. MEG/EEG techniques, non-invasive recordings of macroscopic activities, provide an easy and direct way to examine the system activities which are difficult to be gained and computed in neurophysiology studies.

3.4.2 Information are embedded in endogenous brain oscillations

Considering the novel information MEG/EEG responses provide, new perspectives to examine them are also necessarily required. Different from traditional view regarding MEG/EEG temporal responses as event-related temporal peaks/troughs patterns, it has been suggested for a long time that these event-related potentials are generated by a superposition of evoked oscillations at various frequencies, and in response to stimulus presentation, these brain rhythms undergo a significant resetting or changes in phase and amplitude (Basar, 1998). In other words, this view, distinct from viewing onset MEG/EEG responses (e.g., M100, N1, P1, etc) as some big response peaking from baseline in response to stimulus, interprets the emergence of this peaks/troughs as reorganization and adjustment of ongoing brain rhythms according to incoming stimulus onset (e.g., Makeig et al., 2002; Gruber et al., 2005). This view has several indications: First, brain background activities or spontaneous activities are dominated by endogenous oscillations at various frequencies, different from traditional view of regarding them as a static flat baseline. Secondly, in response to incoming stimulus, these ongoing oscillations modulate their properties (e.g, phase resetting, amplitude resetting, etc) accordingly to track the dynamic structure, different from traditional stimulus-evoked peaks view. Thirdly, the advantage of this view over traditional view is that it can provide a better and realistic brain working mechanism underlying human 'innate world', which is continuously ongoing and not directly related with outside stimulus. This view could also solve a crucial 'context dependent' problem by regarding the background brain oscillations as 'context' in which the incoming information is incorporated and processed. As stated

in a review by Penny et al. (2002), “this is a radically different perspective, which could cast new light onto how cognitive and perceptual processes are implemented in the brain”.

Concerning the tracking mechanism of oscillations, phase and amplitude are the most two important features to describe a change in an oscillation at some specific frequency, and correspondingly, amplitude modulation and phase modulation of these brain rhythms according to outside stimulus are the possible main mechanisms employed. Freeman and Schneider demonstrated the existence of AM mechanism on the olfactory bulb. Specifically, the EEG is a strong periodic waveform, with a spatial distribution of amplitude over the bulb that is consistently different for each specific odor. Partial phase synchronization and resetting has been found to happen in response to visual stimulus onset and working memory task. As described in Chapter 2, our experiment using co-modulated stimuli also confirmed the phase modulation encoding as a way to co-represent the simultaneous envelope and fine structure dynamics.

In this study, we observed the phase modulation of Theta band (4~8Hz) in tracking natural speech sentences, confirming the phase modulation as a general macroscopic activities information carrier. These finding also explained the reasons for the failures in finding the stimulus-specific response in single trials in many previous studies. The observed temporal response in each trial is actually sum of brain oscillations at various frequencies, each of which works in a stimulus related or non-related manner, and therefore, the tracking correlates will be deeply immerged in this summed and complicated oscillation responses. In addition, even when we only

look at responses at theta band, we still could not find the stimulus specific pattern, because it is the phase here rather than amplitude that represents and tracks the incoming stimulus, whereas the temporal waveform is a complex composite of these two factors. These findings also indicate that the main dimensions along which we investigate these macroscopic dynamic responses to seek perceptual correlates are properties of brain oscillation at various frequencies, and it may be one band or multiple bands, and it may be amplitude or phase, or both that contribute to such representation and be of close relevance with behavior. Our previous experiment using simultaneous amplitude- and frequency-modulated sounds also suggest that gamma band (~40 Hz) modulate their phase to track the fine structure dynamics (Luo et al., 2006). Interestingly, in a study investigating somatosensory system of rats (Ahissar et al., 1997), temporally encoded information in sensory input (whisker movements) was found to be decoded actively via 'phase comparators'. Specifically, single units in the somatosensory cortices are found to exhibit spontaneous oscillations around 10 Hz, and the oscillations could track the induced rhythmic whisker movements via the frequency-dependent phase difference. It was further found that these neurons functions as phase comparators that compare cortical timing expectation with the actual timing, rather than passive phase tracking. In sum, tracking stimulus dynamics via phase modulation mechanism is in fact a general mechanism neuron ensembles employ.

3.4.3 200ms temporal processing window

Different cortical oscillations have been found to play different roles in cognitive tasks and here we found it was theta band that shows significant correlates to speech sentence processing and speech intelligibility. The reasons underlying this specific frequency band is probably the statistic temporal structure of speech signals (Dau et al., 1997; Elhilali et al., 2003).

Neurons throughout the auditory system are subject to adaptive processes, tailoring their encoding schemes to match the local sensory environment. Many studies have shown that auditory neurons are tuned for acoustic features found in species-specific communication sounds and process their vocalizations in a more efficient and optimized manner. Neural plasticity studies also found that neurons adjust their response properties according to the statistic structure of incoming complex sounds, to improve the encoding efficiency (Nelken et al., 1999; Dean et al., 2005). Corresponding to the speech processing in human auditory cortex, it is the statistical structure, especially the main dynamic temporal properties of human speech signal (Greenberg, 2003; Poeppel, 2003), that tunes the auditory encoding schemes in human brain.

The basic unit of speech perception is syllable, and reliable information pertaining to syllable appears to be essential for understanding of speeches. The typical length of a syllable in fluent speech is around 200 ms, which is also a important temporal value long enough to provide some measures of perceptual stability through correlations across many parts of the brain, and at the same time, short enough to provide a sufficiently dynamic representation of the stimulus. It is also a ubiquitous interval in

sensory-motor integration, and a minimum time for a motor reaction to occur. In other words, 200 ms is a critical value not only in statistical temporal structure of speech signals, but also reflect a general cortical processing scale. Therefore, it is reasonable that Theta band (4-8 Hz), the frequency corresponding to ~ 200ms, were found to be the main brain working rhythm that track and represent the speech sentence stimuli.

One of the most important roles played by the auditory system is to provide segmental information. It has been shown that hearing impaired patients could gain significant benefits from additional segmental information, which provide independent information about these phonetic boundaries. In absence of such segmentation information, the ability to understand speech will be severely compromised. Therefore, in order to grasp and extract the information of the syllable-like units in the speech stream, the most naïve and efficient brain working mechanism is to analyze the incoming continuous acoustic streams by temporal window with length of also around 200ms. This window slides over time and resets in a way to mark the syllable boundaries, in result, the dynamic speech information is processed and stored in a most efficient manner. This is what exactly observed in this MEG experiment that theta phase pattern could reliable discriminate different speech sentences, which indicates that the 200ms temporal window (period of theta band) resets in a particular pattern (leading to different phase) closely correlated with each specific speech sentence. Note that different speech sentences all contain syllables with same length and rate, but the detailed timing information is distinct. Therefore, to track those different temporal patterns in each speech sentence, the 200ms

temporal window manifest different resetting timing pattern, leading to different phase pattern in theta band observed in our experiment. This interpretation could also account for the noisy classification performances here (see Figure 3.5). If the proposed 200ms temporal window resets according to the syllable boundaries in each sentence, the resulting phase patterns across trials should be coherent only at time points when syllable boundaries occur and not at others. Correspondingly, the classification performance, by using the theta phase pattern of single-trial responses as a template, would deteriorate because of the noisy theta phase values at non-syllable-boundary time points.

A way to test directly whether the theta phase is reset according to syllable boundaries is to look for any relationship between the stimulus and the theta-band phase coherence. This was difficult because there is not a natural way of defining syllable boundaries in terms of the speech spectrogram. I did a preliminary analysis to approximate the stimulus-response relationship, as shown in Figure 3-8. So far, my investigations have not revealed which acoustic features or transients in the speech stimulus are represented or tracked by the theta-band phase in the brain response.

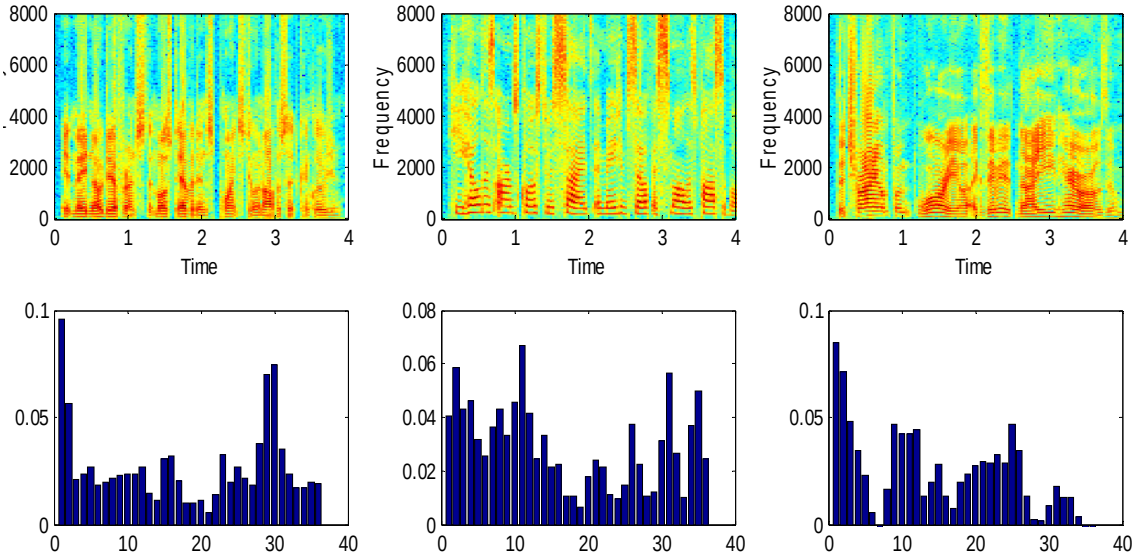


Figure 3-8 Stimulus-response relationships. Upper panel: spectrograms of the 3 natural speech sentences. Lower panel: theta-band phase coherence as a function of time, averaged across 20 channels and 6 subjects.

Another interesting question is that whether these 200ms temporal window found here is a ubiquitous property for all the sensory processing, or is unique for speech signals because of their statistical properties. Further experiments using other auditory signals or in other sensory domain could provide convincing answers.

3.4.4 Stimulus-related or perceptual-related

We also found that the discrimination ability of theta phase pattern in recorded MEG signals were correlated with speech sentence intelligibility. In other words, the phase tracking in theta band was found to be related with behavior and perception. Here we constructed less intelligible speech signals by manipulating their acoustic

properties (e.g., envelop information, fine structure information, etc), and therefore a related interpretation is that maybe such manipulation itself degraded the speech signals, specifically, the boundaries between syllables may become murkier, and may lead to the worse tracking ability of responses. A good test to see whether this tracking is stimulus-related or perceptual-related is to use same stimulus on subjects with different language experiences. For example, non-English subjects may show worse tracking ability when compared to English subjects using same speech sentence stimuli.

3.4.5 Control experiment

We amplitude modulated all of our sentence stimuli at 50 Hz, because we originally expected to find some properties of 50 Hz could track speech sentence dynamics. This hypothesis was motivated by previous experiments (Patel & Balaban, 2000) which showed that when employing amplitude modulated tone sequences, the response phase at the amplitude modulation frequency (37 Hz in that experiment) could track the tone sequence. The observed theta band phase discrimination ability does not depend on the 50 Hz amplitude modulation of the sentences: All stimuli were amplitude modulated at 50 Hz, and the observed discrimination ability was at theta band, far away from the 50 Hz range. We ran a control subject using 4 sentence stimuli (including the 3 used in this experiment) without 50 Hz amplitude modulation and we observed fine classification performance based on the theta band phase pattern and reasonable auditory cortex origin (Figure 3-8)

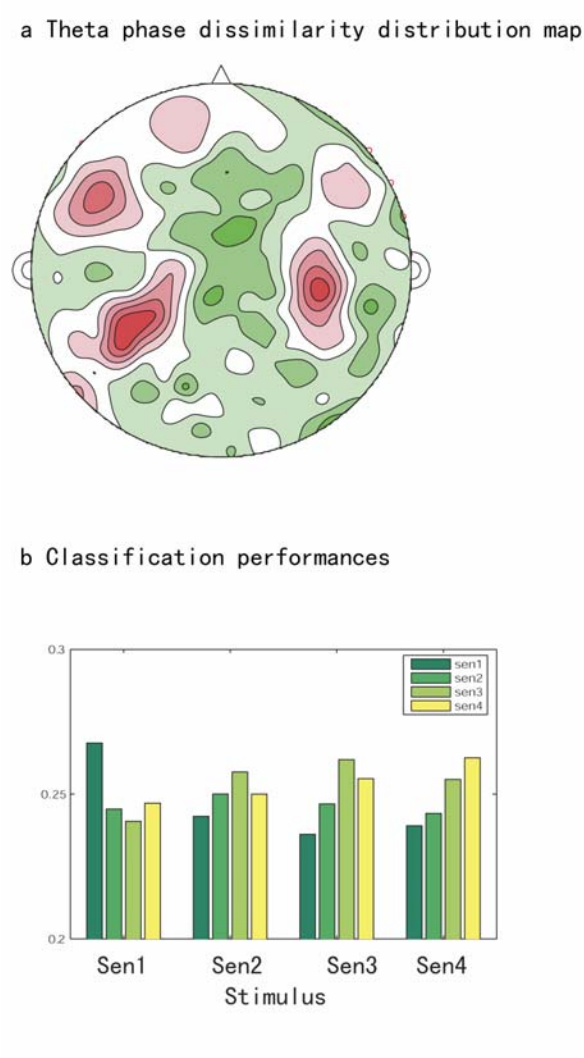


Figure 3-9 Performance of one control subject tested with 4 original speech sentences without amplitude modulation.

3.5 Summary

The coding of natural speech in the auditory cortex remains a central but thorny problem. Neurophysiological studies have shown the remarkable encoding possibilities provided even by single cells. However, these are not linked to speech

intelligibility in humans. Similarly, hemodynamic imaging studies provide data on intelligibility and the role of human auditory cortex, but the recordings are by necessity indirect. In this study, we probed the human auditory system with natural speech sentences and examined the representation mechanism manifested in MEG responses. Specifically, could the MEG response in a single trial contain information that reliably tracks sentences and can be relied on to discriminate sentences? Moreover, we asked whether this discrimination ability is related to speech intelligibility, by additionally employing degraded speech sentences.

We demonstrate that response attributes in single trials of MEG-derived cortical responses suffice to discriminate among sentence stimuli. Specifically, the ongoing phase pattern of theta band (4-8 Hz) responses from human auditory cortex robustly tracks and represents sentences. The discrimination performance evolves over the time of a trial and is strongly present by 1000-2000 ms post-stimulus onset. The ability to distinguish among stimuli is correlated with sentence intelligibility and depends on the difference in both acoustic envelope and fine structure of the speech. Another very novel and consistent finding is that the representation and discrimination ability showed right hemisphere lateralization.

On our view, these data have four major implications. First, the newly observed 'theta phase tracking' mechanism supports the systems neuroscience view of EEG and MEG activity on which such electrophysiological data represent endogenous brain states and reflect stimulus-induced modulation of brain rhythms that are core attributes of the system. Specifically, the ongoing theta band undergoes phase resetting or phase modulation according to temporal transitions in speech sentence

stimuli, leading to the observed discrimination ability of the theta phase pattern. Secondly, the theta band (4~8 Hz) corresponds to a temporal window of 125~250 ms, which is the mean length of syllables across languages, also matching with the sensory efficient coding theory that the auditory system is shaped by the main statistical temporal structure of species-specific communication sounds. Thirdly, the observed rightward lateralization of the theta phase pattern supports a critical role for the right hemisphere in processing speech sentences, specifically at a relatively slower temporal scale of ~200ms (the period of the theta band). Finally, the theta band phase findings can be seen as a direct invasive cortical measure of the speech transmission index (STI), the standard metric used to quantify the relevance of temporal modulation to spoken language understanding in psychophysical research on intelligibility (Greenberg & Arai, 2001; Elhilali et al., 2003).

Cumulatively, our results suggest that sentence stimuli are continuously segmented by a temporal window of ~200 ms duration, a value optimized for one crucial aspect of the statistical temporal structure of speech, roughly the syllable flow. This ongoing sampling window—in our data somewhat biased towards the right hemisphere (see Figure 3.3) even though we are presenting speech—resets in a pattern closely tied to the dynamic structure of speech, including both envelope and fine structure changes. The findings thus support the rightward lateralization of a hypothesized long temporal window in speech and hearing, and the critical role of this temporal window in speech sentence understanding.

Chapter 4: Conclusions

The main goal of the current research is to understand the way in which auditory stimulus dynamics are represented and tracked in the human auditory system, particularly in cortex. Because MEG has been the main tool used in this endeavor, a pertinent question is how magnetic fields recorded from the human scalp should be interpreted and related to the activity of underlying neuron ensembles. This chapter summarizes the results reported in Chapters 2 and Chapter 3 and proposes their implications for general brain mechanisms on the macroscopic scale.

4.1 Research summary

To study the representation of auditory dynamic in human auditory cortex, I probed the human auditory system with a set of stimuli containing rich speech-like temporal dynamic structures and recorded the magnetic responses elicited. In Experiment I, as detailed in Chapter 2, relatively simple auditory stimuli were employed. These stimuli were designed to have dynamics in both amplitude (AM) and fine structure (FM) in order to address a fundamental question in auditory neuroscience: how are these two fundamental temporal modulation features (AM and FM), which always occur simultaneously in natural communication sounds, co-represented in human auditory cortex to achieve 'auditory object unity'? Furthermore, we systematically increased the stimulus dynamics in order to examine possible

encoding transitions concomitant with gradual changes in temporal modulation rate. The key finding in this study is that we identified '*modulation encoding'* as the binding mechanism that unifies these two temporal modulation features. In addition, based on the smooth transition we observed in many aspects of the response, we proposed a corresponding transition from *phase-to-amplitude modulation encoding.*

In Experiment II, detailed in Chapter 3, rather than employing artificially designed 'atomic' dynamic auditory stimuli, natural human speech sentences were used to investigate the neural correlates of natural complex auditory signals. The principal finding was that the *phase of the theta-band* (4~8 Hz) recorded from human auditory cortex robustly tracked speech sentence stimuli in real time, and could be relied on to discriminate sentences, even in single trials. Crucially, this phase tracking ability and robustness strongly depended on the intelligibility of the speech sentence stimuli.

Both experiments addressed the main topic of this thesis-how sound dynamics are tracked by the human auditory system. Importantly, both studies employed either auditory stimuli containing relevant speech-like temporal modulation features or natural speech sentences to investigate the most crucial and challenging issue in cognitive neuroscience-finding the neural correlates of human speech processing and recognition. This issue is also essential to understanding the human auditory system from an evolutionary perspective that proposes the development of its functional structure was geared towards optimizing the processing of species-specific communication sounds. .Furthermore, the two experiments used MEG, an appropriate and advanced non-invasive brain imaging technique with very high temporal

resolution (~1 ms), to monitor brain responses to dynamic auditory stimuli. In the search for the link between the dynamic internal world and the dynamic external world, a new perspective has emerged; researchers are starting to move from past traditional MEG analysis, which focuses on large peaks and troughs in the signal and on dipole localization, and are beginning to explore the spectral domain, motivated by the idea that MEG/EEG signals are really the sum of innate brain oscillations at various frequency bands, each of which plays a distinct and crucial role in sensory and cognitive tasks. Experimental findings based on this new analysis provide novel perspectives on the MEG signal and the relationship between different recording scales (microscopic activity and macroscopic activity). Most importantly, the macroscopic responses reflected in the MEG data were found to be directly pertinent to behavioral perception, suggesting their critical relevance to human cognition.

4.2 Tracking sound dynamics

In single-cell recording studies, neurons have been found to fire spikes that track temporal transients in amplitude or in frequency; these cells also showed selectivity for specific modulations, characterized by the temporal modulation transfer function. At this microscopic level, two main encoding schemes are widely employed to represent and track dynamic sound: *temporal coding* and *rate coding*. The former, by firing spikes temporally locked to stimulus amplitude or frequency transients, explicitly encodes and tracks the ongoing feature changes in an incoming stimulus. The latter, in contrast, implicitly encodes and tracks stimulus dynamics by firing spikes at a rate that is related to the modulation rate of the stimulus. These two

fundamental encoding schemes encompass most if not all of the representation mechanisms reported in neurophyisological studies. This is not really surprising, considering that single cells fire all-or-none spikes with relatively fixed-amplitudes, and the only parameters neurons can adjust in their spike patterns for encoding purposes are spiking rate and spiking time, corresponding to rate coding and temporal coding respectively.

The Experiment I explored the co-representation mechanism for concurrent AM and FM features. The results point to a combinational encoding scheme that simultaneously tracks transients in stimulus amplitude and fine structure. Specifically, this co-representational encoding method unifies rate coding and temporal coding into a combinational encoding scheme by varying the firing rate to represent AM features and by simultaneously varying the local firing time (PM encoding neuron groups) or firing rate (AM encoding neuron groups) to represent FM features.

The Experiment II investigated neural representation of natural speech sentences and revealed a distinct tracking scheme that is robustly present in all 6 human subjects tested and in single trials. The results indicate that it is the ongoing phase of the theta-band (4~8Hz) in the MEG signal that tracks and discriminates speech stimuli, and neither spectral power nor the oscillations at other frequency bands are found to be involved in the representation of speech dynamics. In other words, the stimulus-related dynamic information manifests a more complex format in the corresponding macroscopic activities, and is embedded in and carried by a certain parameter (the phase pattern here) of certain brain oscillation rhythms (here, the theta band). Most interestingly and crucially, this observed tracking ability is correlated

with human subject behavior in that only intelligible speech sentences elicited robust representation; the MEG signal evoked by less intelligible speech shows a less robust tracking pattern, suggesting that this distinct representation and tracking mechanism (embedded in the parameters of certain brain oscillations) found in macroscopic activities may constitute a direct neural correlate of high-level cognitive processing in humans.

Let us return now to Experiment I. If we revisit the results of Experiment I in light of evidence showing the critical role of gamma-band (30~50 Hz) modulations in constructing the aSSR (Ross et al., 2005), we can make the outcome more commensurate with findings from Experiment II by reinterpreting the results of the Experiment I in terms of oscillatory brain activity. Instead of explaining recorded activity from the perspective that neuron groups achieve modulation encoding by manipulating spiking rate/time, we take the position that the recorded signal is the result of oscillatory brain activity and that parameters (amplitude and phase) of the gamma band (30~50 Hz) are modulated to track stimulus dynamics. The gamma band co-represents stimulus transients via modulations in its amplitude and phase.

Unifying Experiment I and Experiment II, we conclude that sound dynamics are tracked by different representational mechanisms and have different formats at the microscopic activity level (single-cell recordings) and the macroscopic activity level (MEG/EEG signals). At the macroscopic level, cortical responses to dynamic stimuli manifest induced, continuous amplitude- or phase- modulation at certain brain rhythms to track the ongoing feature changes in incoming stimuli. Specifically, the gamma band and theta band played critical roles in Experiment I and Experiment II,

respectively The distinct frequency bands observed in these two experiments may be related to the statistical structure of the stimuli and also suggest that the brain may function on multiple temporal scales (temporal period corresponding to frequency band), specifically, a short (~ 40ms) temporal window (corresponding to the gamma band) and a long (~200ms) temporal window (corresponding to the theta band).

4.3 Modulation schemes as a general representation mechanism

Macroscopic brain activity is dominated by oscillations at various frequencies, which are the basic information-carrying components at this level. These rhythms are endogenous innate oscillations; that is, they are continuously present in background brain activities, and they show their own complex dynamics not directly triggered by outside input. As previously discussed, at the single-cell (microscopic) level, rate and time are the two basic parameters of a spike firing pattern that could be employed to carry information, leading to rate coding and temporal coding, respectively. Similarly, at the macroscopic level (reflected in the MEG/EEG signal), where oscillations at various frequencies are the main working components, the amplitude and phase of these oscillations are the two main properties that could be the basic elements of representation, corresponding to amplitude modulation (AM) and phase modulation (PM), respectively (See review by Penny et al., 2002). Both of these modulation representation schemes have been observed in the two experiments, and importantly, they are found to be closely related to perception and cognitive tasks.

Of these two encoding schemes possibly employed by innate brain rhythms, phase modulation (PM) is especially interesting and important because of its capacity to

incorporate more precise temporal information and its closer relationship with the concept of temporal windows. Phase modulation means that the oscillations are constantly resetting their phase to reflect transients in an incoming stimulus, such as stimulus onset, amplitude change, fine structure change, segmental information, etc. In other words, a temporal window with length equal to the period of oscillation rhythms segments the incoming information continuously, resetting the phase in order to track the external transients. The critical roles of the gamma band (30~50 Hz) and the theta band (4~8 Hz) suggest the presence of multiple temporal processing windows in brain working mechanisms, specifically, a short temporal window of ~ 40 ms and a long temporal window of ~ 200 ms.

In sum, the outside world is not represented and reflected in the inside world in a direct way. Instead, these two complex, dynamic systems are coupled indirectly via modulation schemes. Specifically, the inside world works through various innate rhythms that are modulated by the outside world to achieve a unified, dynamic image that bridges the external and internal worlds.

"We are confronted with a system that addresses the external world not as a slumbering machine to be awoken by the entry of sensory information, but rather as a continuously humming brain. This active brain is willing to internalize and incorporate into its intimate activity an image of the external world, but always within the context of its own existence and is own intrinsic electrical activities."

- Rodolfo R. Llinas 'i of the vortex: From Neurons to Self"

4.4 Essence of MEG activities

What new information can be gained from non-invasive brain imaging techniques, specifically MEG? The most obvious and foremost advantage is that MEG allows measurement of human brain activity during normal cognitive tasks, which is impossible to achieve with invasive neurophysiological studies in animals. High temporal resolution and fine spatial resolution are also widely cited benefits of this recording technique. However, even in the ideal case, where we could record from the human brain with high resolution in both the temporal and spatial domains simultaneously, as is possible with electrodes in animal studies, could we really acquire comprehensive knowledge about the mechanisms of the brain? MEG data actually can provide us with novel perspectives and new information about how the brain works.

Typical measured electromagnetic signals require synchronous activation of 10,000-100,000 neurons, and therefore MEG data cannot be treated as the linear sum of single-cell activities, but instead more or less reflect temporal coherence across large neuron populations-in other words, the *'working rhythms'* of the brain machine. Temporal coherence across neurons is believed to be the neurological mechanism that underlies perceptual unity and the conjunction of individually derived sensory components. Therefore, the MEG signal naturally and directly reflects high-level cognitive processing, which is difficult and even impossible to assess from single-cell recordings. In other words, MEG data reflect the system activities and 'brain state' that are closely relevant to perception and behavior, and manifest a dynamic and

flexible coordinates within which external sensory stimuli and internal mental states are unified seamlessly. Correspondingly, the interpretation of the MEG signal should come from the perspective of systems neuroscience: these fluctuating magnetic signals must be understood as modulations of various endogenous, oscillating rhythms in the brain.

"Mapping connectedness in the time domain, superimposed on top of the limited possibilities of spatial connectedness, creates a vastly larger set of possible representations through the almost infinite possibilities of combination".

- Rodolfo R. Llinas 'i of the vortex: From Neurons to Self'

Bibliography

Ahissar E, Haidarliu S, and Zacksenhouse M. Decoding temporally encoded sensory input by cortical oscillations and thalamic phase comparators. Proc Natl Acad Sci U S A 94: 11633-11638, 1997.

Ahissar E, Nagarajan S, Ahissar M, Protopapas A, Mahncke H, and Merzenich MM. Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. Proc Natl Acad Sci U S A 98: 13367-13372, 2001.

Basar E. Brain function and oscillations: Springer, 1998.

Bendor D and Wang X. The neuronal representation of pitch in primate auditory cortex. Nature 436: 1161-1165, 2005.

Bertrand O and Tallon-Baudry C. Oscillatory gamma activity in humans: a possible role for object representation. Int J Psychophysiol 38: 211-223, 2000.

Boemio A. The perceptual representation of acoustic temporal structure: University of Maryland College Park, 2003.

Boemio A, Fromm S, Braun A, and Poeppel D. Hierarchical and asymmetric temporal sensitivity in human auditory cortices. Nat Neurosci 8: 389-395, 2005.

Brosch M, Budinger E, and Scheich H. Stimulus-related gamma oscillations in primate auditory cortex. J Neurophysiol 87: 2715-2725, 2002.

Cariani PA. Temporal codes and computations for sensory representation and scene analysis. IEEE Trans Neural Netw 15: 1100-1111, 2004.

Chait M, Poeppel D, de Cheveigne A, and Simon JZ. Human auditory cortical processing of changes in interaural correlation. J Neurosci 25: 8518-8527,

2005.

Dau T, Kollmeier B, and Kohlrausch A. Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration. J Acoust Soc Am 102: 2906-2919, 1997.

David O, Harrison L, and Friston KJ. Modelling event-related response in the brain. NeuroImage 25: 756-770, 2005.

De Valois RL, Cottaris NP, Mahon LE, Elfar SD, and Wilson JA. Spatial and temporal receptive fields of geniculate and cortical cells and directional selectivity. Vision Res 40: 3685-3702, 2000.

deCharms RC, Blake DT, and Merzenich MM. Optimizing sound features for cortical neurons. Science 280: 1439-1443, 1998.

Dean I, Harper NS, and McAlpine D. Neural population coding of sound level adapts to stimulus statistics. Nat Neurosci 8: 1684-1689, 2005.

Depireux DA, Simon JZ, Klein DJ, and Shamma SA. Spectro-temporal response field characterization with dynamic ripples in ferret primary auditory cortex. Journal of Neurophysiology 85: 1220-1234, 2001.

Dimitrijevic A, John MS, van Roon P, and Picton TW. Human auditory steady-state responses to tones independently modulated in both frequency and amplitude. Ear Hear 22: 100-111, 2001.

Doupe AJ and Kuhl PK. Birdsong and human speech: common themes and mechanisms. Annu Rev Neurosci 22: 567-631, 1999.

Drullman R, Festen JM, and Plomp R. Effect of temporal envelope smearing on speech reception. J Acoust Soc Am 95: 1053-1064, 1994.

Dudley H. The Automatic Synthesis of Speech. Proc Natl Acad Sci U S A 25: 377-383, 1939.

Efron B and Tibshirani RJ. An Introduction to the Bootstrap: Chapman & Hall/CRC, 1994.

Eggermont JJ. Temporal modulation transfer functions for AM and FM stimuli in cat auditory cortex. Effects of carrier type, modulating waveform and intensity. Hear Res 74: 51-66, 1994.

Eggermont JJ and Ponton CW. The neurophysiology of auditory perception: From single units to evoked potentials. Audiol Neuro-Otol 7: 71-99, 2002.

Elhilali M, Chi T, and Shamma SA. A spectro-temporal modulation index (STMI) for assessment of speech intelligibility. Speech Comm 41: 331-348, 2003.

Elhilali M, Fritz JB, Klein DJ, Simon JZ, and Shamma SA. Dynamics of precise spike timing in primary auditory cortex. J Neurosci 24: 1159-1172, 2004.

Engel AK, Fries P, and Singer W. Dynamic predictions: oscillations and synchrony in top-down processing. Nat Rev Neurosci 2: 704-716, 2001.

Fisher NI. Statistical Analysis of Circular Data. Cambridge, UK: Cambridge University Press, 1996.

Friston KJ. Imaging cognitive anatomy. Trends Cogsci 1: 21-27, 1997.

Fitzpatrick DC, Kuwada S, and Batra R. Neural sensitivity to interaural time differences: beyond the Jeffress model. J Neurosci 20: 1605-1615, 2000.

Freeman WJ. Mass Action in the Nervous System. New York: Academic Press, 1975.

Freeman WJ. How brains make up their minds: Columbia Univeristy Press, 2000.

Gaese BH and Ostwald J. Temporal coding of amplitude and frequency modulation in

the rat auditory cortex. Eur J Neurosci 7: 438-450, 1995.

Galambos R, Makeig S, and Talmachoff PJ. A 40-Hz auditory potential recorded from the human scalp. Proc Natl Acad Sci U S A 78: 2643-2647, 1981.

Giraud AL, Lorenzi C, Ashburner J, Wable J, Johnsrude I, Frackowiak R, and Kleinschmidt A. Representation of the temporal envelope of sounds in the human brain. J Neurophysiol 84: 1588-1598, 2000.

Gray CM, Konig P, Engel AK, and Singer W. Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. Nature 338: 334-337, 1989.

Greenberg S and Arai T. The relation between speech intelligibility and the complex modulation spectrum. Eurospeech, 2001.

Greenberg S, Carvey H, Hitchcock L, and Chang S. Temporal properties of spontaneous speech - a syllable-centric perspective. J of Phonetics 31: 465-485, 2003.

Griffiths TD, Warren JD, Scott SK, Nelken I, and King AJ. Cortical processing of complex sound: a way forward? Trends Neurosci 27: 181-185, 2004.

Gruber WR, Klimesch W, Sauseng P, and Doppelmayr M. Alpha phase synchronization predicts P1 and N1 latency and amplitude size. Cereb Cortex 15: 371-377, 2005.

Haller S, Radue EW, Erb M, Grodd W, and Kircher T. Overt sentence production in event-related fMRI. Neuropsychologia 43: 807-814, 2005.

Hamalainen MS. Basic principles of magnetoencephalography. Acta Radiol Suppl 377: 58-62, 1991.

Hamalainen MS. Magnetoencephalography: a tool for functional brain imaging. Brain Topogr 5: 95-102, 1992.

Hancock KE and Delgutte B. A physiologically based model of interaural time difference discrimination. J Neurosci 24: 7110-7117, 2004.

Hari R and Salmelin R. Human cortical oscillations: a neuromagnetic view through the skull. Trends Neurosci 20: 44-49, 1997.

Harms MP and Melcher JR. Sound repetition rate in the human auditory pathway: Representations in the waveshape and amplitude of fMRI activation. Journal of Neurophysiology 88: 1433-1450, 2002.

Hart HC, Hall DA, and Palmer AR. The sound-level-dependent growth in the extent of fMRI activation in Heschl's gyrus is different for low- and high-frequency tones. Hearing Res 179: 104-112, 2003.

Haynes JD and Rees G. Predicting the stream of consciousness from activity in human visual cortex. Curr Biol 15: 1301-1307, 2005.

Haynes JD and Rees G. Predicting the orientation of invisible stimuli from activity in human primary visual cortex. Nat Neurosci 8: 686-691, 2005.

Haynes JD and Rees G. Decoding mental states from brain activity in humans. Nat Rev Neurosci 7: 523-534, 2006.

Heil P. Auditory cortical onset responses revisited .1. First-spike timing. Journal of Neurophysiology 77: 2616-2641, 1997.

Horwitz B, Tagamets MA, and McIntosh AR. Neural modeling, functional brain imaging, and cognition. Trends Cogn Sci 3: 91-98, 1999.

Horwitz B and Poeppel D. How can EEG/MEG and fMRI/PET data be combined?

Hum Brain Mapp 17: 1-3, 2002.

Hsu A, Woolley SM, Fremouw TE, and Theunissen FE. Modulation power and phase spectrum of natural sounds enhance neural encoding performed by single auditory neurons. J Neurosci 24: 9201-9211, 2004.

John MS and Picton TW. Human auditory steady-state responses to amplitude-modulated tones: phase and latency measurements. Hear Res 141: 57-79, 2000.

Kamitani Y and Tong F. Decoding the visual and subjective contents of the human brain. Nat Neurosci 8: 679-685, 2005.

Kawasaki M and Guo YX. Parallel projection of amplitude and phase information from the hindbrain to the midbrain of the African electric fish Gymnarchus niloticus. J Neurosci 18: 7599-7611, 1998.

Klein DJ, Konig P, and Kording KP. Sparse spectrotemporal coding of sounds. Eurasip J Appl Sig P 2003: 659-667, 2003.

Kowalski N, Depireux DA, and Shamma SA. Analysis of dynamic spectra in ferret primary auditory cortex .1. Characteristics of single-unit responses to moving ripple spectra. Journal of Neurophysiology 76: 3503-3523, 1996.

Langers DRM., Backes WH, and van Dijk P. Spectro-temporal Features of the Auditory Cortex: the Activation in Response to Dynamic Ripples. Neuroimage 20: 265-75, 2003.

Lewicki MS. Efficient coding of natural sounds. Nature Neuroscience 5: 356-363, 2002.

Liang L, Lu T, and Wang X. Neural representations of sinusoidal amplitude and

frequency modulations in the primary auditory cortex of awake primates. J Neurophysiol 87: 2237-2261, 2002.

Liegeois-Chauvel C, Lorenzi C, Trebuchon A, Regis J, and Chauvel P. Temporal envelope processing in the human left and right auditory cortices. Cereb Cortex 14: 731-740, 2004.

Llinas R. The intrinsic electrophysiological properties of mammalian neurons: insights into central nervous system function. Science 242: 1654-1664, 1988.

Llinas R and Ribary U. Coherent 40-Hz oscillation characterizes dream state in humans. Proc Natl Acad Sci U S A 90: 2078-2081, 1993.

Llinas R. i of the vortex: from neurons to self. Cambridge, MA: The MIT press, 2000.

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

Luo H, Husain FT, Horwitz B, and Poeppel D. Discrimination and categorization of speech and non-speech sounds in an MEG delayed-match-to-sample study. Neuroimage 28: 59-71, 2005.

Luo H, Wang Y, Poeppel D, and Simon JZ. Concurrent Encoding of Frequency and Amplitude Modulation in Human Auditory Cortex: MEG Evidence. J Neurophysiol, 2006.

Lutkenhoner B and Steinstrater O. High-precision neuromagnetic study of the functional organization of the human auditory cortex. Audiol Neurootol 3: 191-213, 1998.

Machens CK, Gollisch T, Kolesnikova O, and Herz AV. Testing the efficiency of

sensory coding with optimal stimulus ensembles. Neuron 47: 447-456, 2005.

Machens CK, Schutze H, Franz A, Kolesnikova O, Stemmler MB, Ronacher B, and Herz AVM. Single auditory neurons rapidly discriminate conspecific communication signals. Nat Neurosci 6: 341-342, 2003.

Makeig S. Dynamic brain sources of visual evoked responses (vol 295, pg 690, 2002). Science 295: 1466-1466, 2002.

Makeig S. Response: event-related brain dynamics -- unifying brain electrophysiology. Trends Neurosci 25: 390, 2002.

Miller LM, Escabi MA, Read HL, and Schreiner CE. Spectrotemporal receptive fields in the lemniscal auditory thalamus and cortex. Journal of Neurophysiology 87: 516-527, 2002.

Moore BC and Sek A. Detection of frequency modulation at low modulation rates: evidence for a mechanism based on phase locking. J Acoust Soc Am 100: 2320-2331, 1996.

Narayan R, Grana G, and Sen K. Distinct time scales in cortical discrimination of natural sounds in songbirds. J Neurophysiol 96: 252-258, 2006.

Nelken I. Processing of complex stimuli and natural scenes in the auditory cortex. Curr Opin Neurobiol 14: 474-480, 2004.

Nelken I, Rotman Y, and Bar Yosef O. Responses of auditory-cortex neurons to structural features of natural sounds. Nature 397: 154-157, 1999.

Nicolelis MAL, Ghazanfar AA, Faggin BM, Votaw S, and Oliveira LMO. Reconstructing the engram: Simultaneous, multisite, many single neuron recordings. Neuron 18: 529-537, 1997.

Oertel D. Encoding of timing in the brain stem auditory nuclei of vertebrates. Neuron 19: 959-962, 1997.

Oertel D. The role of timing in the brain stem auditory nuclei of vertebrates. Annu Rev Physiol 61: 497-519, 1999.

Ohl FW, Scheich H, and Freeman WJ. Change in pattern of ongoing cortical activity with auditory category learning. Nature 412: 733-736, 2001.

Oppenheim AV and Willsky AS. Signals and systems: Prentice-Hall Inc., 1997.

Papoulis A. The Fourier Integral and its Applications. New York: McGrawHill, 1962.

Patel AD and Balaban E. Temporal patterns of human cortical activity reflect tone sequence structure. Nature 404: 80-84, 2000.

Patel AD and Balaban E. Human auditory cortical dynamics during perception of long acoustic sequences: phase tracking of carrier frequency by the auditory steady-state response. Cereb Cortex 14: 35-46, 2004.

Penfield W and Rasmussen T. The cerebral cortex of man: New York: Macmillan, 1950.

Penny WD, Kiebel SJ, Kilner JM, and Rugg MD. Event-related brain dynamics. Trends Neurosci 25: 387-389, 2002.

Phillips DP, Hall SE, and Boehnke SE. Central auditory onset responses, and temporal asymmetries in auditory perception. Hearing Res 167: 192-205, 2002.

Picton TW, Skinner CR, Champagne SC, Kellett AJ, and Maiste AC. Potentials evoked by the sinusoidal modulation of the amplitude or frequency of a tone. J Acoust Soc Am 82: 165-178, 1987.

Picton TW, John MS, Dimitrijevic A, and Purcell D. Human auditory steady-state responses. Int J Audiol 42: 177-219, 2003.

Poeppel D. The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. Speech Commun 41: 245-255, 2003.

Rees A, Green GG, and Kay RH. Steady-state evoked responses to sinusoidally amplitude-modulated sounds recorded in man. Hear Res 23: 123-133, 1986.

Regan D. Human Brain Electrophysiology: Evoked Potentials and Evoked Magnetic Fields in Science and Medicine. New York: Elsevier, 1989.

Rose HJ and Metherate R. Auditory thalamocortical transmission is reliable and temporally precise. Journal of Neurophysiology 94: 2019-2030, 2005.

Ross B, Borgmann C, Draganova R, Roberts LE, and Pantev C. A high-precision magnetoencephalographic study of human auditory steady-state responses to amplitude-modulated tones. J Acoust Soc Am 108: 679-691, 2000.

Ross B, Herdman AT, and Pantev C. Stimulus induced desynchronization of human auditory 40-Hz steady-state responses. Journal of Neurophysiology 94: 4082-4093, 2005.

Saberi K and Hafter ER. A common neural code for frequency- and amplitude-modulated sounds. Nature 374: 537-539, 1995.

Salinas E and Sejnowski TJ. Correlated neuronal activity and the flow of neural information. Nat Rev Neurosci 2: 539-550, 2001.

Saul AB and Humphrey AL. Spatial and temporal response properties of lagged and nonlagged cells in cat lateral geniculate nucleus. J Neurophysiol 64: 206-224,

1990.

Saul AB, Carras PL, and Humphrey AL. Temporal properties of inputs to direction-selective neurons in monkey V1. J Neurophysiol 94: 282-294, 2005.

Schreiner CE and Urbas JV. Representation of amplitude modulation in the auditory cortex of the cat. I. The anterior auditory field (AAF). Hear Res 21: 227-241, 1986.

Schreiner CE and Urbas JV. Representation of amplitude modulation in the auditory cortex of the cat. II. Comparison between cortical fields. Hear Res 32: 49-63, 1988.

Scott SK, Rosen S, Lang H, and Wise RJ. Neural correlates of intelligibility in speech investigated with noise vocoded speech--a positron emission tomography study. J Acoust Soc Am 120: 1075-1083, 2006.

Seifritz E, Esposito F, Hennel F, Mustovic H, Neuhoff JG, Bilecen D, Tedeschi G, Scheffler K, and Di Salle F. Spatiotemporal pattern of neural processing in the human auditory cortex. Science 297: 1706-1708, 2002.

Sejnowski TJ and Paulsen O. Network oscillations: emerging computational principles. J Neurosci 26: 1673-1676, 2006.

Shannon RV, Zeng FG, Kamath V, Wygonski J, and Ekelid M. Speech recognition with primarily temporal cues. Science 270: 303-304, 1995.

Sharpee TO, Sugihara H, Kurgansky AV, Rebrik SP, Stryker MP, and Miller KD. Adaptive filtering enhances information transmission in visual cortex. Nature 439: 936-942, 2006.

Simon JZ and Wang Y. Fully complex magnetoencephalography. J Neurosci Methods

149: 64-73, 2005.

Singer W and Gray CM. Visual feature integration and the temporal correlation hypothesis. Annu Rev Neurosci 18: 555-586, 1995.

Smith ZM, Delgutte B, and Oxenham AJ. Chimaeric sounds reveal dichotomies in auditory perception. Nature 416: 87-90, 2002.

Smith EC and Lewicki MS. Efficient auditory coding. Nature 439: 978-982, 2006.

Stapells DR, Linden D, Suffield JB, Hamel G, and Picton TW. Human auditory steady state potentials. Ear Hear 5: 105-113, 1984.

Steeneken HJ and Houtgast T. A physical method for measuring speech-transmission quality. J Acoust Soc Am 67: 318-326, 1980.

Suppes P, Han B, Epelboim J, and Lu ZL. Invariance between subjects of brain wave representations of language. Proc Natl Acad Sci U S A 96: 12953-12958, 1999.

Viemeister NF. Temporal modulation transfer functions based upon modulation thresholds. J Acoust Soc Am 66: 1364-1380, 1979.

Wang XQ, Lu T, and Liang L. Cortical processing of temporal modulations. Speech Commun 41: 107-121, 2003.

Williamson S and Kaufman L. Analysis of neuromagnetic signals. In: Handbook of Electroencephalography and Clinical Neurophysiology, edited by Gevins A and Remond A. New York: Esevier: 405-558, 1987.

Wilson MA and Mcnaughton BL. Dynamics of the Hippocampal Ensemble Code for Space. Science 261: 1055-1058, 1993.

Woolley SM, Fremouw TE, Hsu A, and Theunissen FE. Tuning for spectro-temporal

modulations as a mechanism for auditory discrimination of natural sounds. Nat Neurosci 8: 1371-1379, 2005.

Yost WA. Auditory image perception and analysis: the basis for hearing. Hear Res 56: 8-18, 1991.

Zatorre RJ, Belin P, and Penhune VB. Structure and function of auditory cortex: music and speech. Trends Cogn Sci 6: 37-46, 2002.

Zeng FG, Nie K, Stickney GS, Kong YY, Vongphoe M, Bhargave A, Wei C, and Cao K. Speech recognition with amplitude and frequency modulations. Proc Natl Acad Sci U S A 102: 2293-2298, 2005.

Zwicker E. Die Grenzen der Horbarkeit der Amplitudenmodulation und der Frequenzmodulation eines Tones. Acustica 2: 125-133, 1952.