

Neural Representations of Complex Temporal Modulations in the Human Auditory Cortex

Nai Ding¹ and Jonathan Z. Simon^{1,2}

¹Departments of Electrical and Computer Engineering and ²Biology, University of Maryland, College Park, Maryland

Submitted 16 June 2009; accepted in final form 18 August 2009

Ding N, Simon JZ. Neural representations of complex temporal modulations in the human auditory cortex. *J Neurophysiol* 102: 2731–2743, 2009. First published August 19, 2009; doi:10.1152/jn.00523.2009. Natural sounds such as speech contain multiple levels and multiple types of temporal modulations. Because of nonlinearities of the auditory system, however, the neural response to multiple, simultaneous temporal modulations cannot be predicted from the neural responses to single modulations. Here we show the cortical neural representation of an auditory stimulus simultaneously frequency modulated (FM) at a high rate, $f_{FM} \approx 40$ Hz, and amplitude modulation (AM) at a slow rate, $f_{AM} < 15$ Hz. Magnetoencephalography recordings show fast FM and slow AM stimulus features evoke two separate but not independent auditory steady-state responses (aSSR) at f_{FM} and f_{AM} , respectively. The power, rather than phase locking, of the aSSR of both decreases with increasing stimulus f_{AM} . The aSSR at f_{FM} is itself simultaneously amplitude modulated and phase modulated with fundamental frequency f_{AM} , showing that the slow stimulus AM is not only encoded in the neural response at f_{AM} but also encoded in the instantaneous amplitude and phase of the neural response at f_{FM} . Both the amplitude modulation and phase modulation of the aSSR at f_{FM} are most salient for low stimulus f_{AM} but remain observable at the highest tested f_{AM} (13.8 Hz). The instantaneous amplitude of the aSSR at f_{FM} is successfully predicted by a model containing temporal integration on two time scales, ~ 25 and ~ 200 ms, followed by a static compression nonlinearity.

INTRODUCTION

Temporal modulations, including amplitude modulation (AM) and frequency modulation (FM) are important building blocks of complex sounds such as speech and other animal vocalizations. Numerous studies have investigated the neural mechanisms underlying AM and FM processing (Joris et al. 2004). Different neural codes have been found for temporal modulations on different time scales (Bendor and Wang 2007; Wang et al. 2003). Additionally, the neural code of one temporal modulation may be affected by the existence of another temporal modulation on a different time scale (Elhilali et al. 2004). Slow modulations (< 20 Hz) have been most intensively studied because they drive cortical neurons effectively (Eggermont 2002; Liang et al. 2002) and are crucial for speech perception (Drullman et al. 1994; Poeppel 2003; Shannon et al. 1995). At the neuronal level, these slow modulations are encoded by sustained stimulus-synchronized neural activity in cortex. When the stimulus modulation frequency is higher than ~ 20 Hz, individual neurons in primary auditory cortex still precisely phase lock to the stimulus onset but the phase-locking to the periodic features of an AM (or FM) stimuli stops. In other words, slow modulations are represented there

by a temporal code, whereas fast modulations are generally represented by a firing rate code (Bendor and Wang 2007; Wang et al. 2003). These neurons' loss of sustained stimulus-synchronized activity to fast modulations might be caused by temporal integration (Wang et al. 2003) or synaptic depression (Elhilali et al. 2004). In addition to slow modulations, fast modulations of ~ 40 Hz are also well studied, because they evoke strong magnetoencephalography (MEG) and electroencephalography (EEG) responses (Picton et al. 1987; Ross et al. 2000) and are also important for speech perception (Poeppel 2003). Different from single unit responses, the cortical MEG/EEG auditory steady-state responses (aSSR) are phase-locked to the envelope (or carrier frequency fluctuations) of AM (or FM) stimuli with modulation frequency up to ≥ 80 Hz, in a sustained manner (Millman et al. 2009; Picton et al. 1987; Ross et al. 2000). Although the MEG/EEG aSSR phase-locks to the acoustic stimulus features with a precision of a few milliseconds, its temporal dynamics are much slower and show two distinct time scales. When the stimulus modulation frequency is at 40 Hz, the build-up process of aSSR takes > 200 ms, whereas the complementary fall-off process takes < 50 ms (Ross and Pantev 2004; Ross et al. 2002).

Although amplitude and frequency modulations are most commonly used singly in neurophysiology studies, natural sounds always contain multiple amplitude and frequency modulations. For example, speech, the most important acoustic signal for humans, contains at least three levels of modulations (Rosen 1992; Shamma 2006). Each level of modulation has its own contribution to speech perception (Lorenzi et al. 2006; Rosen 1992; Shamma 2006; Shannon et al. 1995; Zeng et al. 2005), and no single level of modulation can extract all the information from speech (Friesen et al. 2001; Remez et al. 1981; Zeng et al. 2005). Although several neural coding strategies have been found for the processing of temporal modulations on different time scales, because of the nonlinearity of the auditory system, it is unlikely that these strategies would be used independently to encode simultaneous multiple levels of temporal modulations (Draganova et al. 2002; Elhilali et al. 2004; Luo et al. 2006, 2007). Hence, it is important to directly determine how these multiple levels of modulations are represented and integrated in the auditory system and which neural computations underlie their processing. The interactions between neural responses to different levels of modulations may show fundamental neural mechanisms in temporal modulation processing. For example, single-unit neural recordings, from primary auditory cortex in ferret, show that the phase-locked neural response to fast temporal modulations is gated by slow temporal modulations (Elhilali et al. 2004). In particular, although most cortical neurons only sustainedly

Address for reprint requests and other correspondence: J. Z. Simon, Department of Electrical and Computer Engineering, University of Maryland, College Park MD 20815 (E-mail: jzsimon@umd.edu).

phase lock to slow AM (<20 Hz), they can sustainably phase lock to ~200-Hz AM if the ~200-Hz AM is further amplitude modulated at a slow rate. If a linear temporal integration model is used to explain this phenomenon, the loss of sustained phase locking to a single AM >20 Hz suggests a 100-ms level long temporal integration window, whereas the precise phase locking to a modulated 200-Hz AM indicates a fine temporal resolution of a few milliseconds. This apparent paradox is explained by a nonlinear model involving synaptic depression (Elhilali et al. 2004).

MEG studies provide further evidence regarding the nonlinear interactions between fast and slow modulations. Luo et al. (2006, 2007) found that when a 37-Hz AM is further frequency modulated at a slower rate f , the MEG response to the 37-Hz AM is primarily phase modulated at f when it is <5 Hz but is simultaneously amplitude and phase modulated at f when it is >8 Hz. Luo et al. (2007) related this transition in MEG response to the transition in FM perception (Moore and Sek 1996) and representations of FM in single neurons (Liang et al. 2002). MEG studies have also shown nonlinear interactions between two fast AMs. Draganova et al. (2002) showed that a pure tone simultaneously amplitude modulated at 38 and 40 Hz elicits MEG responses not only at 38 and 40 Hz but also at 2 Hz, which could be related to the perceptual beating frequency of the stimulus. Ross et al. (2003) showed that, for monaural stimuli, the MEG response to one auditory AM is suppressed by another AM in the stimulus. Fujiki et al. (2002) found the MEG response to a monaural AM stimulus is also suppressed by another AM stimulus sent to the contralateral ear. In contrast, EEG studies (Dimitrijevic et al. 2001; John et al. 2001) have shown that the neural responses to simultaneous fast FM (40–80 Hz) and fast AM (40–80 Hz) of the same carrier are relatively independent. In sum, all these MEG/EEG studies agreed, when multiple temporal modulations exist in the stimulus, each of them would evoke an aSSR at its modulation frequency. In addition, neural responses are also observable at frequencies like the summation or difference of two stimulus modulation frequencies, because of the interaction between two stimulus modulations. These interaction terms provide a possible neural correlate of FM or beating rhythm perception (Draganova et al. 2002; Luo et al. 2007).

In this work, we show the neural interactions between responses to fast FM (near 40 Hz) and slow AM (<15 Hz) of the same carrier, using MEG. Both fast FM and slow AM are extremely important ingredients of speech. Formant transitions in speech are frequency fluctuations on a time scale of 20–40 ms (Poeppel 2003), corresponding to 25–50 Hz, whereas the envelope of speech (Rosen 1992) is an amplitude fluctuation on a time scale of 150–300 ms (Drullman et al. 1994; Greenberg et al. 2003; Poeppel 2003), corresponding to 3–6 Hz. Therefore understanding the neural coding of simultaneous slow AM and fast FM can provide us insights into the neural coding of speech. In addition, although there has not yet been a MEG study on 40-Hz FM, the EEG aSSR evoked by 40-Hz FM is known to be strong and have similar temporal dynamics with MEG aSSR evoked by 40-Hz AM (Poulsen et al. 2007). Therefore by parametrically amplitude modulating the 40-Hz FM, we can also study how the two time scales, seen separately in the response build-up and fall-off of 40-Hz aSSR (Ross and Pantev 2004), manifest themselves together in the response to a stimulus continuously changing its amplitude.

Based on the discussions above, for stimuli containing simultaneous fast FM (near 40 Hz) and slow AM (<15 Hz), we propose to examine 1) whether both the fast FM and slow AM features are represented by precise temporal coding in the auditory cortex, reflected by phase-locked aSSR at both f_{FM} and f_{AM} ; 2) whether the power, phase, and phase locking of the neural responses change as a function of the stimulus f_{AM} ; 3) whether the aSSR at f_{FM} is additionally amplitude and/or phase modulated with fundamental frequency f_{AM} , analogous to the effect seen in earlier studies using simultaneous fast AM and slow FM (Luo et al. 2006, 2007); and 4) whether the temporal characteristics of the aSSR to fast modulations near 40 Hz found in the current and previous studies (Poulsen et al. 2007; Ross and Pantev 2004; Ross et al. 2002) can be explained by a unified computational mechanism.

METHODS

Subjects

Ten young adult subjects (18–34 yr old; mean age, 24.6 yr; 5 females) participated in this study. One additional subject participated but was not included in the analysis because of overly strong magnetic artifacts caused by dental work. All subjects reported normal hearing and no history of neurological disorder. The experimental procedures were approved by the University of Maryland institutional review board. Written informed consent was obtained from each subject before the experiment. All subjects were paid for their participation.

Stimuli

The stimulus in this study is a tone that has been simultaneously frequency modulated and amplitude modulated. Mathematically, this “FM-AM” stimulus is given by $x(t) = [1 - \cos(2\pi f_{AM}t)]\cos[2\pi f_C t + \Theta(t)]$, where $\Theta(t) = (\Delta f/f_{FM})\sin(2\pi f_{FM}t)$. The carrier frequency, f_C , is fixed at 550 Hz, consistent with previous MEG studies using related sounds (Luo et al. 2006, 2007). The FM modulation frequency, f_{FM} , is fixed at 37.7 Hz. The maximum frequency deviation from the carrier frequency, Δf , is 330 Hz. The AM modulation frequency, f_{AM} , is varied, taking on values of 0.3, 0.7, 1.7, 3.1, 4.9, 9.0, and 13.8 Hz. The specific values of f_{FM} and f_{AM} are selected to avoid harmonic overlaps. One example waveform, and its instantaneous frequency, is plotted in Fig. 1. The seven total stimulus conditions are differentiated only by their f_{AM} . Stimulus duration is 21 s, including 100 ms rising

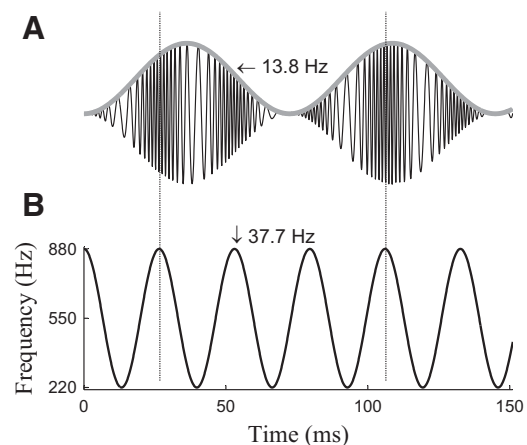


FIG. 1. A: the waveform of the FM-AM stimulus with f_{AM} equal to 13.8 Hz. The envelope of the waveform, marked by a gray line, is oscillating at 13.8 Hz. B: instantaneous frequency of the stimulus shown in A. The instantaneous frequency is modulated at 37.7 Hz for every stimulus in this study.

and falling cosine ramps. The stimuli were generated digitally using MATLAB (The MathWorks, Natick, MA), with sampling frequency of 44.1 kHz. All stimuli have the same root mean square value and therefore have similar perceptual loudness (Moore et al. 1999). To ensure subjects' attention to the stimuli throughout the entire duration of the experiment, we also included two oddball stimuli for every stimulus condition, for which subjects were instructed to press a button in response to the presentation of the oddball. In the first oddball stimulus, f_{FM} is momentarily doubled for a 180-ms portion of the stimulus. In the other, the envelope of stimulus was momentarily made constant for 140 ms for a portion of the stimulus. In this way, listening for the features of the oddball stimuli encourages subjects to attend to both the fast FM and slow AM features in the stimulus. Neural responses to oddball stimuli are not analyzed.

During the MEG recordings, stimuli were delivered using tube-telephone (E-ARTONE 3A 50 X, Etymotic Research) at a comfortable loudness of ~ 75 dB SPL. Standard and oddball stimuli are presented in pseudo-random order, with interstimulus interval randomly chosen from between 1.8 and 2.2 s. Each individual standard stimulus is repeated eight times, whereas each individual oddball stimulus is played once. In sum, during the experiment, 56 (7 stimulus conditions \times 8 trials) standard stimuli and 14 (7 stimulus conditions \times 2 types) oddball stimuli are presented. The task given to the subjects is to press a button whenever they hear an oddball stimulus. The false alarm rate and miss rate averaged over all 10 subjects are both 4%. The duration of the entire experiment is ~ 40 min, divided into five blocks; subjects may rest between blocks.

MEG recordings and denoising

The neuromagnetic signals were recorded using a 157-channel whole-head MEG system (5-cm baseline axial gradiometer SQUID-based sensors, KIT, Kanazawa, Japan). Three reference channels are also used to measure environmental magnetic field. All recordings are conducted in a magnetically shielded room, sampled at 1 kHz. A 200-Hz low-pass filter and a notch filter at 60 Hz are applied on-line. Two denoising techniques were applied off-line: time shifted principle component analysis (TS-PCA) (de Cheveigné and Simon 2007), which removes external noise (filtered versions of the reference channel signals), and sensor noise suppression (SNS) (de Cheveigné and Simon 2008b), which removes noise arising internally from individual gradiometers. TS-PCA was used with a ± 100 ms range of time-shifts (filter taps); SNS was used with 10 channel-neighbors to exclude sensor noise. Finally, DSS (de Cheveigné and Simon 2008a), a blind source separation technique designed to preserve phase-locked neural activities, was applied. The DSS components are sorted based on how much percent of the response power is phase-locked to the stimulus. Only the first DSS component was kept for further analysis in this study.

aSSR source localization

The neural source of the aSSR is localized with respect to the source of the N1m response, a prominent peak ~ 100 ms after the onset of a pure tone, also called the M100 (Näätänen and Picton 1987). To localize the aSSR's neural source, all subjects participated in a preliminary experiment, listening to 100 repetitions of a 1-kHz pure tone (duration, 50 ms; interstimulus interval randomized between 750 and 1,550 ms). A transient auditory evoked response to these short tones was observable in the MEG recording of every subject. The source of the N1m and aSSR in each hemisphere is localized using an equivalent-current dipole model in the MEG Laboratory software program v.2.001M (Yokogawa Electric, Eagle Technology, Kanazawa Institute of Technology). Only cortical sources are considered because MEG is insensitive to subcortical neural sources.

Analysis in the frequency domain

The MEG response to each standard stimulus (21-s duration) is extracted as a single trial. For each subject, there are eight trials of each stimulus condition. Within a few hundred milliseconds after each stimulus onset, the transient auditory-evoked response diminishes and the aSSR builds up (Ross et al. 2002). Because this study focuses on steady-state responses, the first second of the MEG response in each trial is removed from analysis. The last 20-s recording of each trial is transformed into the frequency domain using Fourier analysis (with 0.05-Hz resolution), allowing responses at f_{FM} and f_{AM} to be easily resolved. After Fourier analysis, the MEG response at any given frequency is represented by a complex number possessing both magnitude and phase. The square of its magnitude is called the power at that frequency. The response power for a particular stimulus condition is defined as the power of the MEG response averaged over all eight trials of that stimulus condition. The power of this average (in contrast to the average of the power) is dominated by phase-locked power and is called evoked power. The evoked power is affected by both the power in a single trial and the phase coherence over trials. The signal-to-noise ratio at any frequency is defined as the ratio between the power at that frequency and the power averaged over a 0.1-Hz range on both sides.

Sometimes, we are interested in the relation between the response power and the stimulus frequency rather than the absolute response power. In this case, we reduce subject-to-subject variability by normalizing the response power of each subject by dividing the response power at a specific frequency (e.g., f_{AM}) for each stimulus condition by the mean response power at that frequency averaged over all stimulus conditions. If the response power is measured in decibels, this normalization is equivalent to subtracting the mean response power across stimuli at that frequency from the response power at that frequency for that particular stimulus condition. This normalization thus preserves the relation between the response power and the stimulus frequency for every single subject. To reduce the influence of background noise, we estimate the power of the neural response as the difference between the power of the MEG response and the power of the background noise. It is motivated by the assumption that the MEG response at a particular frequency is a linear combination of the stimulus specific response and background, stimulus irrelevant, neural activity that is uncorrelated with the stimulus specific response.

Extraction of amplitude and phase modulation

In this study, the aSSR at f_{FM} is hypothesized to be amplitude modulated and/or phase modulated at f_{AM} , in which case the neural response to fast FM can be modeled as

$$y(t) = [A + E(t)]\cos[2\pi f_{FM}t + \Phi(t)] \quad (1)$$

where $E(t)$ and $\Phi(t)$ are both periodic with fundamental frequency f_{AM} , $A + E(t)$ is always positive, and $E(t)$ has zero mean. Using the Fourier transform, the aSSR at f_{FM} is decomposed into a number of sinusoids with constant amplitude and phase. That is, although the aSSR at f_{FM} can be represented by a single sinusoid at f_{FM} with periodically time varying amplitude and phase, its Fourier decomposition shows that it can also be represented as the sum of multiple sinusoids at f_{FM} (the "carrier" frequency), and at $f_{FM} \pm f_{AM}$, $f_{FM} \pm 2f_{AM}$, etc. (the "sideband" frequencies). The modulation functions $E(t)$ and $\Phi(t)$ determine the amplitude and phase of each Fourier component but cannot be inferred directly from any single component. It is relatively easy to determine whether $E(t)$ or $\Phi(t)$ is a constant, however, based on a statistic called the α parameter (Luo et al. 2006, 2007). The α parameter is defined as $\alpha = 2\angle Y(f_{FM}) - \angle Y(f_{FM} - f_{AM}) - \angle Y(f_{FM} + f_{AM})$, where $Y(f)$ is the Fourier transform of $y(t)$ at frequency f (Luo et al. 2006). The α parameter is π when $E(t)$ is a constant but is 0 when $\Phi(t)$ is a constant. If the α parameter is neither

π nor 0, neither $E(t)$ nor $\Phi(t)$ is a constant, indicating the aSSR at f_{FM} is modulated in both amplitude and phase.

Different from the Fourier transform, which removes all time information and replaces it with frequency information, the Hilbert transform keeps all time information and can be used to extract the instantaneous amplitude and phase from any signal directly. The instantaneous amplitude extracted by the Hilbert transform is also called the Hilbert envelope. Because the derivative of instantaneous phase is instantaneous frequency, modulating the phase of a signal is equivalent to modulating the frequency of that signal. In this paper, the term FM will only be used to describe acoustic signals, whereas the term phase modulation will only be used to describe neural signals, consistent with the conventions of the different fields (Luo et al. 2006, 2007; Moore 2003). Based on Eq. 1, the Hilbert instantaneous amplitude and phase of the aSSR at f_{FM} are $A + E(t)$ and $2\pi f_{FM}t + \Phi(t)$, respectively. The Hilbert instantaneous phase $2\pi f_{FM}t + \Phi(t)$ contains information about both f_{FM} and the phase modulation function $\Phi(t)$. Because we are only interested in the phase modulation function $\Phi(t)$, which can be obtained by subtracting the $2\pi f_{FM}t$ from the Hilbert instantaneous phase, we will refer to $\Phi(t)$ as the instantaneous phase in the remainder of the paper. The constant (DC) component of the instantaneous amplitude, A , reflects the power rather than the temporal dynamics of the aSSR at f_{FM} . Hence it will be called the power of the aSSR at f_{FM} and can be estimated by the power of MEG response at f_{FM} . The fluctuating (AC) component of the instantaneous amplitude, $E(t)$, determines all the temporal dynamics of the aSSR. Its power (i.e., its mean square) will be called the power of the instantaneous amplitude. Because $E(t)$ is assumed to be periodic with fundamental frequency f_{AM} , its power only arises from frequencies at f_{AM} , $2f_{AM}$, $3f_{AM}$, etc. Experimentally, the $E(t)$ extracted by the Hilbert transform is often noisy; therefore to reduce the influence of noise, the power of the instantaneous amplitude is estimated as the sum of the powers of the instantaneous amplitude at f_{AM} , $2f_{AM}$, $3f_{AM}$, etc.

When the aSSR at f_{FM} is exactly sinusoidally amplitude modulated, its instantaneous amplitude is $A + K \sin(2\pi f_{AM}t + \varphi)$, where the modulation depth is K/A . In this special case, K can be estimated based on the power of the instantaneous amplitude. When the aSSR at f_{FM} is amplitude modulated by a more general waveform with fundamental frequency f_{AM} , $E(t) = A + Ke(t)$, where $e(t)$ is a function with unit maximum amplitude and zero mean. In this general case, K can only be estimated from the dynamic range of $E(t)$, but we will still refer to the ratio K/A as a modulation depth. To avoid possible confusion with the modulation depth of the acoustic stimulus, we will call the ratio K/A the neural AM modulation depth.

The above discussion shows that the Hilbert transform is a suitable tool to analyze the aSSR at f_{FM} ; however, it cannot be applied to the MEG response directly because the MEG response also contains the aSSR at f_{AM} and additional background noise. Thus before doing a Hilbert transform, we filter the MEG response between 20.7 and 54.7 Hz. This filtering process removes both the aSSR at f_{AM} and background noise <20.7 Hz or >54.7 Hz but preserves the MEG response at f_{FM} and $f_{FM} \pm f_{AM}$ in every stimulus condition.

Confusion matrix analysis

In this study, we are interested in whether the neural response at a particular frequency $f_a^{\text{neural}}(f_{AM})$ is tracking the acoustic feature f_{AM} . The frequencies $f_a^{\text{neural}}(f_{AM})$ could be, for example, f_{AM} , $f_{FM} + f_{AM}$, or $2f_{AM}$ for $a = 1, 2, 3, \dots$. If the neural response at one frequency, e.g., $f_2^{\text{neural}}(f_{AM}) = f_{FM} + f_{AM}$ is observable, its power should be significantly larger than zero. Consequently, the MEG power at that frequency, $PP_a(f_{AM}) = |Y(f_a^{\text{neural}}(f_{AM}))|^2$, should be significantly larger than the power of background noise at $f_a^{\text{neural}}(f_{AM})$. The seven stimulus f_{AM} values, $f_{AM}^1, f_{AM}^2, \dots, f_{AM}^7$ (i.e. f_{AM}^m for $m = 1, 2, \dots, 7$) in the current experiment are selected so that the frequency of most neural responses of interest, e.g., $f_a^{\text{neural}}(f_{AM}) = f_{AM}, 2f_{AM}, 3f_{AM}$,

$f_{FM} \pm f_{AM}, f_{FM} \pm 2f_{AM}$, or $f_{FM} \pm 3f_{AM}$, is unique to one specific stimulus f_{AM} value. For example, if $f_a^{\text{neural}}(f_{AM}^1)$ is a possible neural response frequency when the stimulus f_{AM} is f_{AM}^1 then it is not a possible neural response frequency when the stimulus f_{AM} is $f_{AM}^2, f_{AM}^3, \dots$ or f_{AM}^7 . Consequently, for a specific stimulus f_{AM} value, e.g., f_{AM}^1 , not only can the power of the neural component of the MEG response at $f_a^{\text{neural}}(f_{AM}^1)$ be measured by the power at that frequency, for the condition with stimulus f_{AM} given by f_{AM}^1 , but also the power of the background noise at the frequency $f_a^{\text{neural}}(f_{AM}^1)$ can be estimated by the power at the same frequency, averaged over all six stimulus conditions whose stimulus f_{AM} is not f_{AM}^1 . Because the MEG response measured at frequency $f_a^{\text{neural}}(f_{AM}^m)$, $m = 1, 2, \dots, 7$, depends on the stimulus condition $n = 1, 2, \dots, 7$, it can be more generally represented as $P_a(f_{AM}^m, f_{AM}^n) = |Y(f_a^{\text{neural}}(f_{AM}^m))|^2|_{f_{stim}=f_{AM}^n}$. $P_a(f_{AM}^m, f_{AM}^n)$ naturally forms a 7×7 matrix, called the confusion matrix of neural response $f_a^{\text{neural}}(f_{AM})$. Each row of the confusion matrix contains MEG responses at the same frequency, whereas each column of the confusion matrix contains a MEG response from the same stimulus condition. The ratio between the power in a diagonal element, $P_a(f_{AM}^m, f_{AM}^m)$, and the power $P_a(f_{AM}^m, f_{AM}^n)$ averaged over all off-diagonal elements on the same row ($n \neq m$), is a signal-to-noise ratio. Furthermore, the ratio is also an element of a statistical $F(2, 12)$ distribution (John and Picton 2000) and will be used to test the statistical significance of the neural signal.

A model of the instantaneous amplitude of the aSSR near 40 Hz

The amplitude of the aSSR is sensitive to the intensity of acoustic stimulus (Ménard et al. 2008; Ross et al. 2000). Hence changes in stimulus intensity will be reflected in the instantaneous amplitude of aSSR. Here, we describe temporal changes of the stimulus intensity as an acoustic gating function. For example, when the stimulus is a 40-Hz FM turned on for 300 ms and then turned off, the acoustic gating function is a 300-ms-long rectangular window, equal to the Hilbert envelope of the stimulus. However, the acoustic gating function is not always the Hilbert envelope. When the stimulus is a 40-Hz AM turned on for 300 ms and turned off, the acoustic gating function is still a 300-ms-long rectangular window, whereas the Hilbert envelope will be a product of the acoustic gating function and a 40-Hz sinusoid resulted from the 40-Hz AM. In this case, the acoustic gating function is the envelope of the Hilbert envelope, sometimes called the envelope (Ewert et al. 2002). The acoustic gating function describes intensity changes in stimuli (in dB) and will serve as the input of the model of the instantaneous amplitude of an aSSR near 40 Hz.

The temporal dynamics of an aSSR near 40 Hz are well studied using rectangular windows as acoustic gating functions (Poulsen et al. 2007; Ross and Pantev 2004; Ross et al. 2002). As stated in the Introduction, the build-up and fall-off processes of the 40-Hz aSSR occur with two distinct scales. In addition, the aSSR is also sensitive to extremely brief change in the stimulus. Even a 3-ms temporal gap in the stimulus modulation can cause a temporary reduction in the amplitude of aSSR. After the gap, the aSSR builds up similarly to its build-up at the onset of a stimulus. The slow build-up process indicates a temporal integration window >200 ms, whereas the fast fall-off process indicates a temporal integration window <50 ms. The asymmetry of the build-up process and fall-off process indicates the system generating 40-Hz aSSR is nonlinear. Partly because of this nonlinearity, the build-up and fall-off time constant of aSSR is different from its apparent latency (Ross et al. 2000). Based on these experimental observations, we proposed a system using two temporal integration constants and a compression nonlinearity to explain the temporal dynamics of the amplitude of the aSSR near 40 Hz. If the acoustic gating function is denoted as $x(t)$, the instantaneous aSSR amplitude is denoted as $r(t)$, and the amplitude compression function is denoted as $\Gamma(\cdot)$, the model can be expressed as

$$\begin{aligned}
 \tau_{\text{fast}} \frac{dr_1(t)}{dt} &= -r_1(t) + \lambda x(t) \\
 \tau_{\text{slow}} \frac{dr_2(t)}{dt} &= -r_2(t) + \lambda x(t) \\
 r(t) &= \Gamma(r_2(t) + r_1(t))
 \end{aligned}
 \quad (2)$$

where $r_1(t)$ and $r_2(t)$ are the outputs of the fast and slow temporal integration systems, respectively, whereas λ is a constant factor determining the magnitude of the aSSR evoked by a given stimulus. To optimally fit the MEG recording from Ross and Pantev (2004), the two time constants, τ_{fast} and τ_{slow} , are chosen to be 25 and 200 ms, whereas the amplitude compression function $\Gamma(\cdot)$ is chosen to be a hard-thresholding function that is equal to its input when positive but 0 otherwise. Because this model involves a linear temporal integration process and a nonlinear amplitude compression process, it is called an “integration-compression” model. This model is extremely simple but can fit MEG recordings in the literature (Ross and Pantev 2004; Ross et al. 2002) very well. In particular, when driven by the acoustic gating function of the stimulus used by Ross and Pantev (2004), the model outputs the waveform in Fig. 2, which compares well with the MEG data of the same study. In the simulation, the aSSR builds up slowly and falls off fast. Because of the fast fall-off, the aSSR can respond to a temporal gap as short as 3 ms. This simulation shows the integration-compression model is successful when the acoustic gating function represents a binary state, on or off. The question whether this model can be applied for continuously varying acoustic gating function is addressed below, using the stimulus in this study. Because this study uses near 40-Hz FM, the acoustic gating function is just the slow AM in the stimulus.

RESULTS

The spatial power distribution of the MEG response of a representative subject is shown in Fig. 3A. The location of the neural source of aSSR is 10 mm more medial, on average, than the source of N1m (95% CI, 2–18 mm), but not significantly different in other directions. The localization result is consistent with previous MEG aSSR studies (Ross et al. 2005). The power spectral densities of MEG responses to three representative stimuli are shown in Fig. 3B. We will first report the MEG responses at exactly f_{AM} and f_{FM} and then focus on the temporal dynamics of the aSSR at f_{FM} . When the aSSR with center frequency f_{FM} is amplitude and/or phase modulated, its power will spread around f_{FM} . In this paper, the term “aSSR at f_{FM} ” or “40-Hz aSSR” will include all the temporal dynamics of an aSSR whose center frequency is f_{FM} or near 40 Hz (37.7 Hz in our case) and therefore will refer to a changeable narrowband signal rather than a pure, constant amplitude, sinusoid.

MEG response at f_{AM}

An MEG response at the stimulus f_{AM} is observed in all stimulus conditions. The confusion matrix of the power of the MEG response at f_{AM} is shown in Fig. 4A. The power of the MEG response at f_{AM} is significantly [$F(2,12) > 7.1$, $P < 0.01$] larger than the power of background noise at f_{AM} in all tested stimulus conditions, as shown by the confusion matrix analysis. The relation between the strength of neural AM response and the stimulus f_{AM} is commonly known as modulation transfer function (MTF) (Ross et al. 2000; Viemeister 1979). The MTF measured by the power of the MEG response at f_{AM} has a low-pass pattern: the power of the MEG response to an AM sound decreases with increasing f_{AM} of that sound (visible in the diagonal elements of Fig. 4A). However, because the power of background noise also shows an approximate $1/f$ power spectral density (Fig. 3B) and the MEG response at f_{AM} is a mixture of both stimulus-driven aSSR at f_{AM} and background noise, it needs to be clarified whether the low-pass pattern of the MTF results from stimulus-driven aSSR or background noise. With the assumption that the MEG response is a linear combination of stimulus-driven aSSR and independent background noise, one can estimate the power of stimulus-driven aSSR by subtracting the estimated power of background noise at f_{AM} from the power of measured MEG signal at f_{AM} (see METHODS). The MTF measured by this corrected power of the aSSR at f_{AM} still shows a low-pass pattern and can be modeled as a linear function of f_{AM} measured in Hz (Fig. 4B). The slope of the fitted linear function is -0.96 dB/Hz (99% CI, -1.17 to -0.74 dB/Hz). For $f_{\text{AM}} > 1$ Hz, the slope of the MTF can also be fitted as -3.6 dB/oct (99% CI, -4.8 to -2.5 dB/oct). Because the slope of the fitted line is significantly negative ($P < 0.01$), the low-pass pattern of the MTF is statistically significant for $f_{\text{AM}} < 15$ Hz. This low-pass pattern is also seen in surface EEG (Picton et al. 1987) and intracranial EEG (Liegeois-Chauvel et al. 2004). In the regression analysis, 7 of 70 (7 stimulus conditions \times 10 subjects) available samples are removed because their power is not statistically significant ($P > 0.05$). To reduce subject-to-subject variability, the corrected power is normalized (see METHODS) before the regression analysis. Even without any correction, normalization, and sample rejection, the slope of the MTF is still significantly negative (99% CI, -1.73 to -0.29 dB/Hz). To study whether the reduction in the evoked power of the MEG response at f_{AM} is caused by a loss of energy in every single

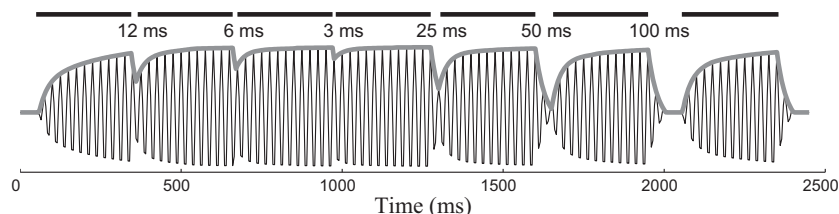


FIG. 2. The 40-Hz auditory steady-state response (aSSR) in a simulated gap-detection experiment described by Ross and Pantev (2004). The simulation uses the integration-compression model described in this paper. The acoustic stimulus in this stimulation contains 7 bursts of 300-ms-long 40-Hz AM tones. The gaps (intervals) between the 7 bursts of sound are 12, 6, 3, 25, 50 and 100 ms, respectively. The black bars on the top of the figure show when a burst of 40-Hz AM tone is present. The thick gray curve is the simulation result of the integration-compression model: the instantaneous amplitude of the 40-Hz aSSR. The fast-changing thin black line under the thick gray line is the simulated 40-Hz aSSR, which is obtained by multiplying a 40-Hz sinusoid with the simulated instantaneous amplitude. Experimental magnetoencephalography (MEG) recordings for the gap detection experiment can be found in Fig. 3 in Ross and Pantev (2004).

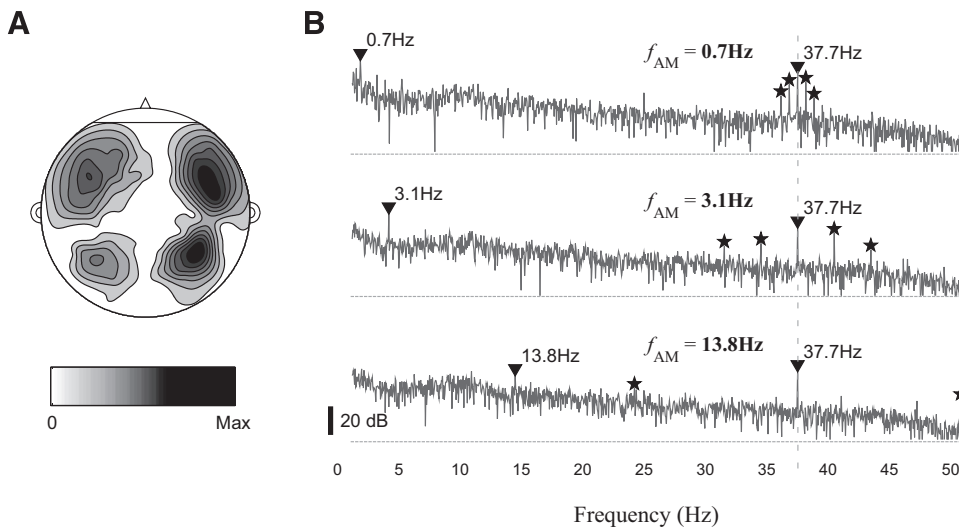


FIG. 3. MEG recordings from a representative subject (R1170). *A*: the spatial distribution of the power of MEG recordings on a flattened head. The grayscale map used here will be applied to all the other figures in the paper. *B*: power spectral densities of MEG responses to 3 representative stimuli. The f_{AM} of the stimulus are 0.7, 3.1, and 13.8 Hz for the top, middle, and bottom panels, respectively. The MEG responses at f_{AM} and f_{FM} are marked by inverted triangles, whereas the responses at $f_{AM} \pm f_{AM}$ or $f_{FM} \pm 2 f_{AM}$ are marked by stars.

trial or a loss of phase locking over trials, we calculated the phase coherence value (Fisher 1993) of the MEG response at f_{AM} over trials. For the seven tested f_{AM} , from 0.3 to 13.8 Hz, the grand averaged phase coherence values of aSSR at f_{AM} are 0.76, 0.95, 0.90, 0.96, 0.93, 0.91, and 0.85, respectively. One-way ANOVA shows the phase coherence values does not significantly change when the stimulus f_{AM} increases from 0.7 to 13.8 Hz [$F(5,54) = 0.84$, $P > 0.5$]. Hence, the low-pass pattern of evoked power of aSSR at f_{AM} is caused by a change in single trial power rather than a change in over trial phase coherence. Because both the neural response power and background noise power are strongest at low frequencies, regression analysis was used to show that the signal to noise ratio of the neural response at f_{AM} does not significantly increase or decrease when f_{AM} increases ($P > 0.6$). Confusion matrix analysis also shows the power of the MEG response at $2f_{AM}$ is statistically significant except for $f_{AM} = 1.7$ or 4.9 Hz [$F(2,12) > 4.0$, $P < 0.05$], whereas the power of the MEG response at $3f_{AM}$ is statistically significant for $f_{AM} > 1.7$ Hz [$F(2,12) > 4.2$, $P < 0.05$]. If the neural response power at f_{AM} , $2f_{AM}$, and $3f_{AM}$ are combined as the neural response power to the stimulus AM, the MTF has a slope of -1.06 dB/Hz (99% CI, -1.30 to -0.82 dB/Hz).

MEG response at f_{FM}

In addition to the MEG response at the stimulus f_{AM} , a strong MEG response at f_{FM} is also observed (see Fig. 3*B*). The

signal-to-noise ratio of the MEG response at f_{FM} is >15 dB in every stimulus condition for every subject. This result agrees with previous EEG results (John et al. 2001; Picton et al. 1987) that tones frequency modulated near 40 Hz can evoke strong aSSR. The power of the MEG response at f_{FM} decreases almost linearly with the stimulus f_{AM} measured in octaves (Fig. 5*A*). Linear regression shows the slope of the fitted linear function is -0.86 dB/oct (99% CI, -1.06 to -0.62 dB/oct). Because the slope is significantly negative ($P < 0.01$), the power of the MEG response at f_{FM} significantly decreases with the stimulus f_{AM} , at a rate around 0.86 dB/oct. In the regression analysis, the response power is normalized (see METHODS), but even without this normalization, the slope of the fitted line is still significantly negative ($P < 0.01$). In contrast to the response power, as is shown in Fig. 5*B*, the grand averaged phase coherence value of the MEG response at f_{FM} is >0.9 in every stimulus condition and does not significantly interact with the stimulus f_{AM} [1-way ANOVA, $F(6,63) = 0.68$, $P > 0.65$]. Hence, increasing f_{AM} reduces the power rather than the phase locking ability of the MEG response at f_{FM} . Because all stimuli have the same intensity, the change in aSSR power must result from the temporal dynamics of aSSR. The relation between the power of the MEG response at f_{FM} and the stimulus f_{AM} is successfully predicted by the integration-compression model (Fig. 5*A*). The correlation coefficient between the MEG response power predicted by the integration-compression model and that obtained by experimental data grand

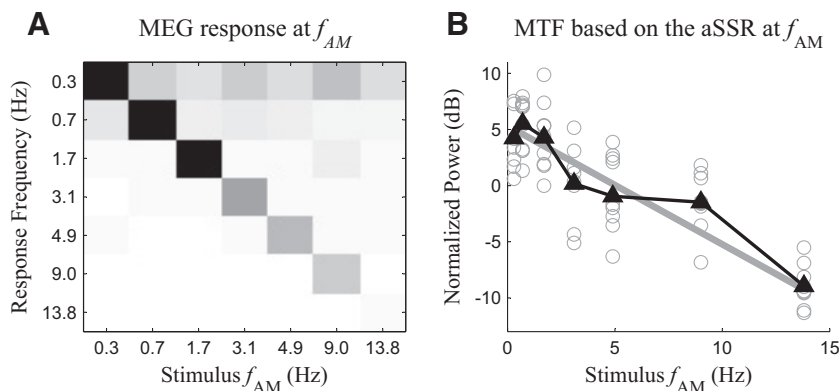


FIG. 4. Analysis of the power of the aSSR at f_{AM} . *A*: confusion matrix of the power of the MEG response at f_{AM} . The diagonal elements of the confusion matrix are the power of aSSR at f_{AM} , without any normalization or correction. The grayscale map is as shown in Fig. 3. *B*: the modulation transfer function (MTF) calculated as a function of the corrected power of the aSSR at f_{AM} and stimulus f_{AM} . Each gray hollow circle represents the corrected power of the aSSR at f_{AM} for one subject. The black line marked by triangles shows the grand averaged MTF. The gray line is the optimal linear fit of the MTF.

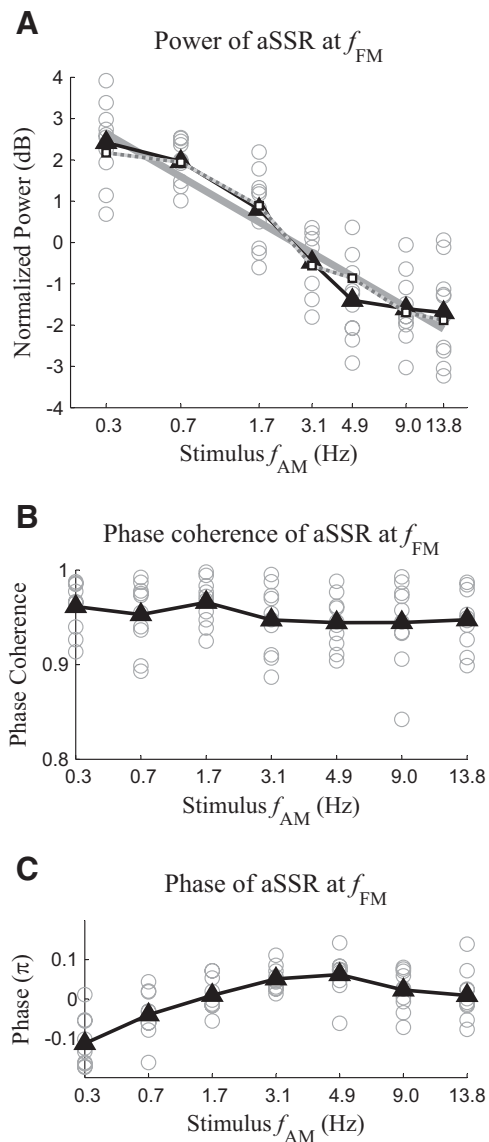


FIG. 5. Power (A), phase coherence over trials (B), and phase (C) of the MEG response at f_{FM} . Each gray hollow circle represents the data collected from one subject. The solid black curve marked by triangles shows the grand average over subjects. The solid gray line in A is the optimal linear fit of the relation between the normalized power in decibels and f_{AM} in octaves. The dotted gray line with white square markers in A shows the result predicted by the compression-integration model.

averaged over all the subjects is 0.99. Not only the power but also the phase of the MEG response at f_{FM} significantly interacts with the stimulus f_{AM} [Fig. 5C, 1-way ANOVA, $F(6,63) = 13.02$, $P < 0.01$]. To reduce subject-to-subject variability, for each subject, the mean phase averaged over stimulus conditions is subtracted from the phase in each stimulus condition. Even without this normalization, one-way ANOVA still shows a significant interaction between the phase of the MEG response at f_{FM} and the stimulus f_{AM} [1-way ANOVA, $F(6,63) = 2.66$, $P < 0.03$].

Temporal dynamics of the aSSR at f_{FM}

One of the primary goals of this work is to examine the interaction between fast modulations and slow modulations.

One piece of evidence for such an interaction is the existence of MEG responses at sideband frequencies such as $f_{FM} \pm f_{AM}$ and $f_{FM} \pm 2f_{AM}$ (Fig. 3B). Because acoustic stimuli in this study only contain temporal modulations at f_{FM} and f_{AM} , responses at these combinational modulation frequencies around f_{FM} cannot be explained as a direct tracking of acoustic modulations. Confusion matrices for the power of the upper ($f_{FM} + f_{AM}$) and lower ($f_{FM} - f_{AM}$) sidebands are shown in Fig. 6, A and B. An F -test shows both sidebands are statistically significant [$F(2,12) > 19.4$, $P < 0.001$ for upper sideband and $F(2, 12) > 7.5$, $P < 0.01$ for lower sideband] in every stimulus condition. Similarly the MEG response at $f_{FM} + 2f_{AM}$ is statistically significant [$F(2, 12) > 10.25$, $P < 0.01$] when f_{AM} is < 9 Hz, and the response at $f_{FM} - 2f_{AM}$ is statistically significant [$F(2,12) > 4.9$, $P < 0.03$] at $f_{AM} = 0.3, 0.7, 3.1$, and 4.9 Hz. These sidebands indicate the aSSR at f_{FM} is temporally modulated at f_{AM} (Luo et al. 2006, 2007). To determine whether it is the amplitude or phase of the aSSR at f_{FM} that is modulated at f_{AM} , we calculate the α parameter, which is a phase relationship between the MEG responses at f_{FM} and $f_{FM} \pm f_{AM}$ (Luo et al. 2006, 2007); α is 0 for pure amplitude modulation and π for pure phase modulation. In this case, however, α falls between 0 and π (Fig. 6C), which indicates the modulation is a mixture of amplitude modulation and phase modulation. To confirm the co-existence of amplitude modulation and phase modulation, we analyze the instantaneous amplitude and phase of the aSSR at f_{FM} using the Hilbert transform. Both the instantaneous amplitude and phase show strong periodicity with fundamental frequency at f_{AM} . Confusion matrix analysis (Fig. 7, A and B) shows that, for both the instantaneous amplitude and phase of the aSSR at f_{FM} , the response power at f_{AM} is significantly stronger [$F(2,12) > 7.9$, $P < 0.01$ for instantaneous amplitude and $F(2,12) > 13.6$, $P < 0.001$ for instantaneous phase] than the background noise power at f_{AM} , in every stimulus condition (the power of the instantaneous phase has no simple interpretation and is used only as a statistic to detect periodicity). In sum, the α parameter analysis and the instantaneous amplitude and phase analysis both shown that the aSSR at f_{FM} is simultaneously amplitude and phase modulated with fundamental frequency f_{AM} .

Because the instantaneous amplitude of the aSSR at f_{FM} oscillates with fundamental frequency f_{AM} , it is a neural correlate of the stimulus slow AM. Consequently, the relation between the power of the instantaneous amplitude and the stimulus f_{AM} can also be regarded as an effective MTF. To the extent that the aSSR at f_{FM} is sinusoidally amplitude modulated, the power of its instantaneous amplitude concentrates at f_{AM} . With this assumption, the MTF has a slope of -0.59 dB/Hz (99% CI, -0.92 to -0.28 dB/Hz). If the aSSR at f_{FM} is not sinusoidally amplitude modulated, its power will be distributed at f_{AM} and its harmonics. For this possibility, we estimate the power of the instantaneous amplitude as the sum of the power at the first four harmonics of f_{AM} (the 5th and higher-order harmonics are not considered because they are not statistically significant in any stimulus condition). The MTF calculated this way has a slope of -0.72 dB/Hz (99% CI, -0.95 to -0.49 dB/Hz). When f_{AM} is > 1 Hz, the slope of the MTF can also be fitted as -3.0 dB/oct (99% CI, -5.0 to -1.6 dB/oct). For this MTF calculation, the power at each harmonic of f_{AM} is corrected by subtracting the power of background noise at the corresponding frequency; the estimate of the power

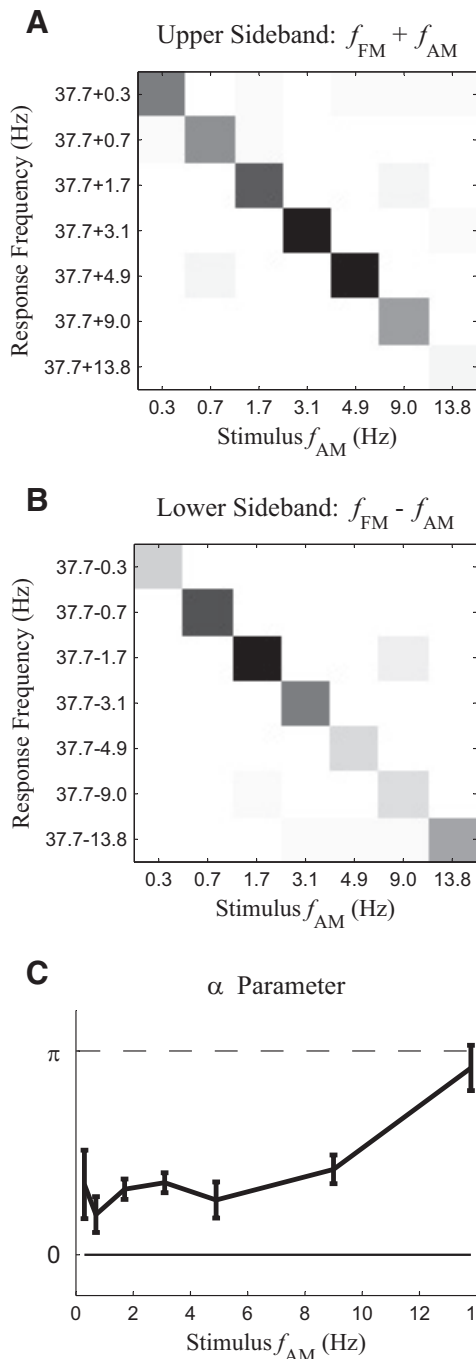


FIG. 6. Analysis of MEG responses at $f_{FM} + f_{AM}$ (upper sideband) and $f_{FM} - f_{AM}$ (lower sideband). *A* and *B*: confusion matrices for the power of the upper sideband (*A*) and lower sideband (*B*). The grayscale map is as shown in Fig. 3. (Note: in *B*, the response at 37.7–1.7 Hz from $f_{AM} = 9$ Hz may be enhanced because of being the 4th harmonic of 9 Hz. In that case, the estimate of background noise would be conservatively high but would not affect the statistical significance of any of the diagonal responses.) *C*: the α parameter for each stimulus condition. If the aSSR is only modulated in its amplitude, the α parameter should be 0, marked by the solid line. If the aSSR is only modulated in its phase, the α parameter should be 1, marked by the dashed line. Error bars indicate circular SD (Fisher 1993).

of the instantaneous amplitude is also normalized to reduce subject-to-subject variability (see METHODS for details regarding the subtractive correction and normalization). Samples not statistically significant ($P > 0.05$) are excluded from the linear

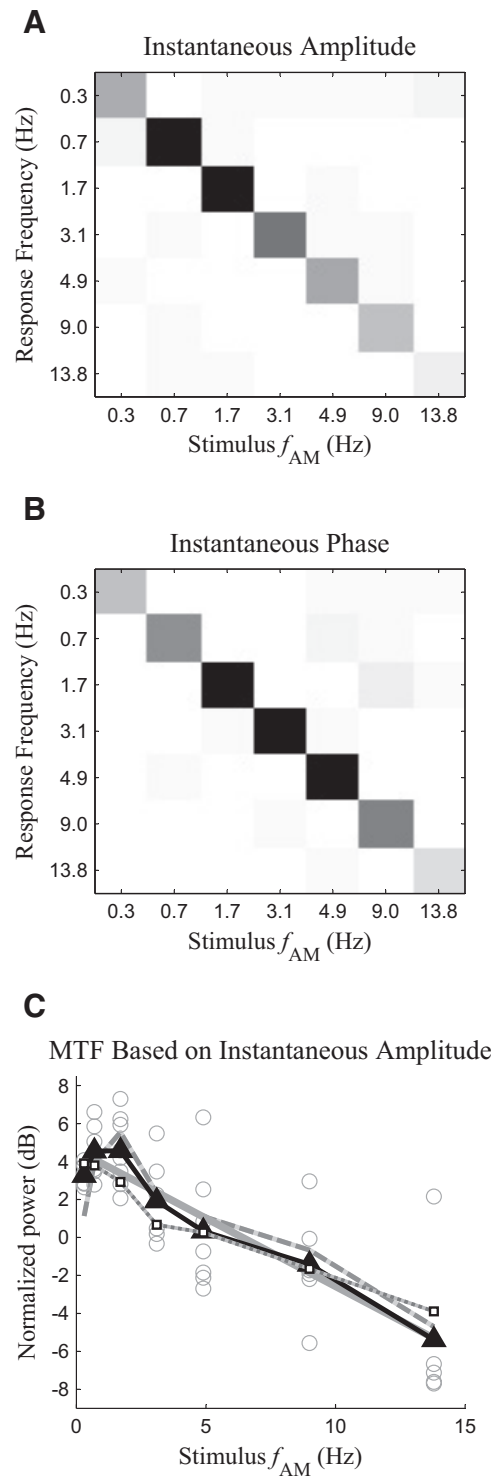


FIG. 7. Analysis of the instantaneous amplitude and instantaneous phase of the aSSR at f_{FM} . *A* and *B*: confusion matrices of the power of the instantaneous amplitude (*A*) and phase (*B*) at f_{AM} . The grayscale map is shown in Fig. 3. *C*: MTF calculated based on the power of the instantaneous amplitude of the aSSR at f_{FM} . The solid black line with triangle markers is the MTF averaged over all subjects. The solid gray line is the optimal linear fit of the MTF, whereas the dotted gray line with white square markers is the MTF predicted by the integration-compression model. The power of the instantaneous amplitude for each subject and each condition is shown as a gray hollow circle. The instantaneous amplitude's power at f_{AM} is plotted as the dashed gray line.

regression analysis. Even without any correction, normalization, and sample rejection, the slope of the MTF is still significantly negative ($P < 0.01$) whether the aSSR at f_{FM} is assumed to be sinusoidally amplitude modulated or not. The low-pass pattern of the MTF measured from the instantaneous amplitude is also successfully predicted by the integration-compression model (Fig. 7C). The correlation coefficient between the MTF predicted by the integration-compression model and that obtained by experimental data grand averaged over all the subjects is 0.97.

As the stimulus f_{AM} increases, the power of the MEG response at f_{FM} decreases at 0.86 dB/oct, whereas the power of the instantaneous amplitude of the aSSR at f_{FM} decreases 3.0 dB/oct. As discussed in METHODS, if the aSSR at f_{FM} is assumed to be sinusoidally amplitude modulated, the neural AM modulation depth of the aSSR can be estimated based on the ratio between the power of the instantaneous amplitude of the aSSR and the power of the MEG response at f_{FM} . Hence, with the sinusoidal AM assumption, the neural AM modulation depth should decrease at 2.1 dB/oct. Without the sinusoidal AM assumption, the neural AM modulation depth can only be estimated from the waveform of the instantaneous amplitude. We plot the grand averaged waveforms of the instantaneous amplitude and phase of the aSSR at f_{FM} over one stimulus AM period in Fig. 8. The waveforms of the instantaneous amplitude and phase are roughly in phase. Both the instantaneous amplitude and phase change most dramatically when the intensity of the acoustic stimulus is near its minimum. The dynamic range of the instantaneous amplitude of the aSSR at f_{FM} decreases

with increasing stimulus f_{AM} , indicating that the neural AM modulation depth is decreasing. Linear regression analysis shows that the dynamic range of the instantaneous amplitude decreases with the stimulus f_{AM} at 1.2 dB/oct (99% CI, 1.0–1.4 dB/oct). In contrast to the dynamic range of the instantaneous amplitude, when the stimulus f_{AM} is ≤ 3.1 Hz, the dynamic range of the instantaneous phase is almost constant ($\sim 0.5\pi$), corresponding to a latency shift of 6 ms for a 40-Hz signal, the same value as seen at the onset or reset point of 40-Hz aSSR (Ross et al. 2005). As the stimulus f_{AM} increases beyond 4.9 Hz, the dynamic range of the instantaneous phase begins to decrease.

DISCUSSION

This study showed strong and predictable interactions between the neural representations of simultaneous fast FM and slow AM of the same carrier, consistent with the predictions of an integration-compression model. In particular, both the instantaneous amplitude and phase of the aSSR at f_{FM} oscillate with fundamental frequency f_{AM} (but with an interaction pattern quite different from that found when the stimulus contains simultaneous slow FM and fast AM of the same carrier; Luo et al. 2006). Because of these neural interactions, the information in slow AM is simultaneously encoded in neural oscillations at f_{AM} and f_{FM} . As discussed in the following text, recent psychoacoustical studies on complex temporal modulations suggest both neural representations of the slow AM are important for auditory perception (Fullgrabe and Lorenzi 2005; Lorenzi et al. 2001a). In this experiment, both the neural activities at high (37.7 Hz) and low (< 14 Hz) frequencies are purely stimulus driven. It is worth mentioning that the power of spontaneous gamma band oscillations (30–50 Hz) is also found to be modulated by low frequency neural oscillations (< 10 Hz) in macaque monkey auditory cortex (Lakatos et al. 2005). It is unclear whether the purely stimulus driven coupling between the neural responses at different modulation rates found in this and previous studies (Elhilali et al. 2004; Luo et al. 2006, 2007) is related to this spontaneous coupling between different frequency bands. The MTF measured by the power of the aSSR at f_{AM} has a low-pass pattern, consistent with previous neurophysiological and psychoacoustical studies, as discussed in the following text. With the increase of f_{AM} , the power and the neural AM modulation depth of the aSSR at f_{FM} also decrease, well predicted by the integration-compression model. The success of the integration-compression model is evidence for dual time constants in the auditory system.

Neural encoding of AM and FM

FM and AM are of great importance in speech perception and processing of natural sounds in general. However, how they are processed by the auditory system is still not well understood (Joris et al. 2004), including debates over whether FM and AM are processed using the same or different mechanisms (Eggermont 2002; Gaese and Ostwald 1995; Hart et al. 2003; Liang et al. 2002; Picton et al. 1987; Saberi and Hafter 1995). Psychoacoustical studies indicate that dichotic FM and AM stimuli can be perceived as coming from a single intracranial imaginary location (Saberi and Hafter 1995). Extracellular recordings indicate that the best FM modula-

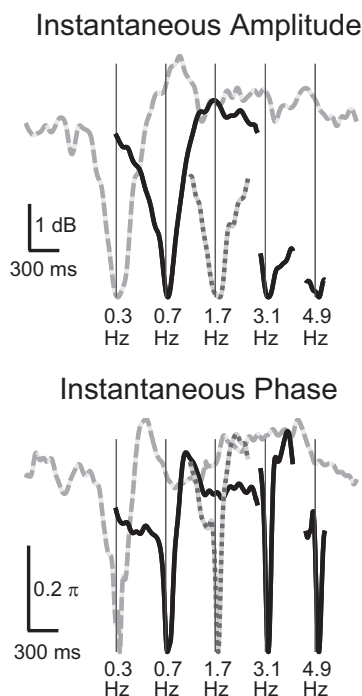


FIG. 8. Waveforms of the instantaneous amplitude (top) and phase (bottom) of the aSSR at f_{FM} during 1 AM period of the stimulus. The stimulus f_{AM} is labeled at the bottom of each curve. To facilitate comparison, all waveforms are aligned based on their minimum value. Each curve is shifted by 500 ms relative to the curve on its right to reduce overlap. A vertical thin line shows when the acoustical input in each stimulus condition reaches its minimum. Waveforms are not plotted when the stimulus f_{AM} is ≥ 9 Hz because their periods are too short to be seen clearly on the current time scale.

tion frequency and best AM modulation frequency of individual neurons in primary auditory cortex are highly correlated (Eggermont 2002; Gaese and Ostwald 1995; Liang et al. 2002). Functional MRI (fMRI) provides evidence that FM and AM sounds activate a common cortical area (Hart et al. 2003). However, in primary auditory cortex, there are also neurons sensitive to FM but not to AM (Eggermont 2002; Gaese and Ostwald 1995) and there are neurons sensitive to the direction and speed of FM (Eggermont 2002; Gaese and Ostwald 1995). This study, and previous MEG FM-AM studies (Luo et al. 2006), provide further evidence that FM and AM components are represented with distinct neural codes. If the neural codes for FM and AM are the same, the MEG response to simultaneous fast FM and slow AM should be the same as that to simultaneous fast AM and slow FM. However, we found the aSSR to fast FM is simultaneously amplitude and phased modulated by the slow AM, whereas Luo et al. (2006) found the aSSR to fast AM is primarily phase modulated by the slow FM (with a transition to single sideband modulation for faster FM rates). The α parameter of MEG responses to simultaneous fast FM and slow AM is near π (Luo et al. 2006), whereas α for simultaneous fast AM and slow FM is significantly different from π and is even closer to 0. The asymmetry in cortical MEG response to FM-AM sound indicates FM and AM have distinct neural codes, independently of whether they are processed by the same neural population.

Neurophysiological (Liang et al. 2002; Luo et al. 2007) and psychoacoustical studies (Moore and Sek 1996) also showed that the neural encoding and perception of slow modulations are different from those of fast modulations. One cannot yet conclude whether the asymmetric MEG response to FM-AM sound results from the slow FM/AM or the fast FM/AM or both. Based on available neurophysiological and psychoacoustical results, however, it seems the asymmetry may depend more on slow FM/AM. With a 550-Hz carrier frequency, slow FM (<5 Hz) is perceived as a fluctuation in pitch, whereas slow AM (<5 Hz) is perceived as a fluctuation in loudness. The modulation detection thresholds for slow FM and AM are also very different (Moore 2003). However, both 40-Hz FM tones and 40-Hz AM tones are both perceived as "harsh or rough" (Joris et al. 2004; Picton et al. 1987) and have similar modulation detection threshold (Moore 2003). The aSSR to 40-Hz FM and AM are also reported to be similar (John et al. 2001; Picton et al. 1987).

Modulation transfer function

The MTF describes neural or perceptual sensitivity to temporal modulations. The temporal modulated stimuli used in MTF experiments vary from click trains (Eggermont 2002), amplitude modulated white noise, or ripple noise (Chi et al. 1999; Viemeister 1979) to amplitude/frequency modulated tones (Eggermont 2002; Liang et al. 2002). The neural sensitivity or perceptual sensitivity is measured by either the firing pattern of a single neuron (Eggermont 2002; Liang et al. 2002), the power of MEG response (Ross et al. 2000), hemodynamic response (Giraud et al. 2000), or psychoacoustical modulation detection threshold (Chi et al. 1999; Moore 2003; Viemeister 1979). Evidence from single-unit (Eggermont 2002; Liang et al. 2002), local field potentials (Eggermont 2002; Liegeois-

Chauvel et al. 2004), fMRI (Giraud et al. 2000), and psychoacoustical studies (Chi et al. 1999; Moore 2003; Viemeister 1979) all agree the MTF has a low-pass pattern when the stimulus modulation frequency is higher than ~ 2 Hz, independently of the kind of temporally modulated stimulus. However, MEG and EEG studies are inconsistent about the shape of the MTF. The prevalence of $\sim 1/f$ low frequency background in MEG/EEG measurements makes it difficult to detect and estimate the power of low-frequency aSSR. Hence, the focus of many MEG/EEG studies has been on fast FM/AM at 40 Hz or even higher (Dimitrijevic et al. 2001; John et al. 2001; Picton et al. 1987; Ross and Pantev 2004; Ross et al. 2000, 2002, 2005), rather than slow FM/AM. One EEG study (Picton et al. 1987) reported aSSR to slow FM/AM (<15 Hz) is measurable and, when statistically significant, is most salient at 2–5 Hz, indicating that the MTF from 2 to 15 Hz may have a low-pass pattern. This study also shows that the aSSR to slow AM (<15 Hz) is measurable (at least in the presence of simultaneous fast FM) and decreases in its power when the modulation frequency increases. The MTF measured by the power of the instantaneous amplitude of the aSSR at f_{FM} also shows a low-pass pattern for $f_{AM} < 15$ Hz. The slopes of the both MTF are between 3 and 4 dB/oct. We know of no psychoacoustical study measuring the MTF using a fast FM carrier. However, the psychoacoustical MTF measured with a fast AM carrier has a slope between 3 and 4 dB/oct in the low-frequency range (<15 Hz), consistent with our neurophysiological results (Lorenzi et al. 2001b).

Multiple representations of slow modulations

Slow temporal modulations (<15 Hz) are crucial for speech intelligibility. In quiet, slow temporal modulations alone are enough for speech recognition (Drullman et al. 1994; Shannon et al. 1995). However, removing fast modulations in speech makes speech intelligibility more susceptible to noise (Friesen et al. 2001). Hence, the importance of fast modulations of speech, i.e., fine structure, is also emphasized in studies of speech perception (Lorenzi et al. 2006; Zeng et al. 2005). However, the neural mechanisms underlying the importance of speech fine structure are still not well known. This study shows that the presence of fast modulations can facilitate the neural representation of slow modulations. Without the fast modulations, a slow modulation is only represented by the neural oscillation directly tracking its modulation frequency. However, with fast modulations, the slow modulation is also represented by the instantaneous amplitude and instantaneous phase of the neural response to fast modulations. In other words, the presence of fast modulations adds redundancy to the representation of slow modulations. If this conclusion can be generalized to general complex sounds containing simultaneous fast modulations and slow modulations, perhaps the neural responses to fast modulations in speech, such as periodicity oscillations and formant transitions (Poeppel 2003; Rosen 1992; Shamma 2006), can make the neural representation of slow modulations in speech more robust, by giving additional cues about the slow modulations.

Although this study shows the presence of multiple neural representations of slow modulations in the auditory cortex, it does not provide evidence about whether these multiple neural representations contribute to sound perception. Recent psy-

choacoustical studies of second-order AM (Fullgrabe and Lorenzi 2005; Lorenzi et al. 2001a), however, show that the perception of slow modulations in a complex sound does depend on multiple mechanisms. Second-order AM sound contains two levels of modulations. The carrier signal is first amplitude modulated at a relatively high-frequency f_1 , e.g., 60 Hz, and then further amplitude modulated at a slow frequency f_2 , e.g., 3 Hz. The slow temporal modulation at f_2 is called the second-order AM of the carrier. Generalizing the result of this study, one may hypothesize that the information of the second-order AM will be simultaneously encoded in neural oscillations at f_2 and temporal dynamics of neural oscillations at f_1 . If both neural codes of the second-order AM can be used by the nervous system, interferences at a single modulation frequency, e.g., f_2 , can affect but cannot completely abolish subjects' sensitivity to the second-order AM. If only one neural code, e.g., the neural oscillation at f_2 , can be decoded by the nervous system, subjects' sensitivity will be completely eliminated by interferences at one modulation frequency, e.g., f_2 , and will not be affected by interferences at the other modulation frequency, e.g., f_1 . Psychoacoustical results support the idea that both neural representations contribute to the perception of second-order AM, because the detection threshold of the second-order AM is degraded but is not completely abolished when intrinsic fluctuations at low modulation frequencies, around f_2 , are introduced (Lorenzi et al. 2001a). More evidence that the second-order modulation information can be decoded from both the neural oscillations at f_1 and f_2 comes from a study on the perceptual beating frequency of the second-order AM (Fullgrabe and Lorenzi 2005). When f_1 is <20 Hz, the perceptual beating frequency of the second-order AM depends primarily on the time interval between the main peaks of its f_1 component, whereas when f_1 is >20 Hz, the perceived beating frequency depends primarily on f_2 . Moreover, the same study also showed that different subjects may depend on different representations to decode the slow modulation information. For example, as is shown in Fig. 3 of that work, one subject (H.C.) estimates the beating frequency of the second-order AM as the time interval between the main peaks of f_1 component even when f_1 is ~ 40 Hz, whereas the two other subjects (M.H. and P.E.) estimate the beating frequency as f_2 even when f_1 is <20 Hz.

Temporal integration and temporal dynamics of 40-Hz aSSR

Temporal integration is believed to be an important property of the auditory system (Moore 2003). However, time constants of the temporal integration window measured from different psychoacoustical experiments seem inconsistent. Loudness perception and modulation detection experiments show a temporal integration window >200 ms, whereas gap detection experiments show a fine temporal resolution of a few milliseconds (Moore 2003; Ross et al. 2002). A similar paradox exists in speech perception experiments; humans are sensitive to fast formant transition cues (20–40 ms) but can also do a temporal integration over a few hundred milliseconds to extract syllabic level information (Poeppel 2003). The 40-Hz aSSR is a good tool to study the neurophysiology underlying the “resolution-integration paradox” (de Boer 1985) for three reasons. First, the build-up of 40-Hz aSSR takes >200 ms (Ross and Pantev 2004; Ross et al. 2002), which may correspond to the long

200-ms level temporal integration window. Second, the post-stimulus fall-off of 40-Hz aSSR occurs within 50 ms (Ross and Pantev 2004; Ross et al. 2002), implying an integration time <50 ms. Third, the 40-Hz aSSR can be interrupted by a temporal gap as short as 3 ms, showing the extremely fine resolution of the auditory system. In sum, the temporal dynamics of 40-Hz aSSR reflects both the slow integration and fine temporal resolution properties of the auditory system. As is shown by the successful fits of the integration-compression model, the “resolution-integration paradox” in the temporal dynamics of 40-Hz aSSR can be well explained by a simple nonlinear system with two time constants.

It is constructive to compare more closely the integration-compression model with models addressing the psychoacoustical “resolution-integration paradox” (de Boer 1985). One type of psychoacoustical model suggests dual temporal windows exist in the auditory system and are implemented by different but overlapped neural populations (Poeppel 2003). This model is supported by recent fMRI studies (Boemio et al. 2005; Giraud et al. 2007). The integration-compression model is similar in nature to this model. It is interesting that the integration-compression model's two time constants, 25 and 200 ms, obtained by fitting the temporal dynamics of 40-Hz aSSR, coincide with the two time constants suggested by Poeppel (2003) based on speech perception experiments. However, because the two temporal integration systems in the integration-compression model are both linear and operate strictly in parallel, they can factually be unified as a single system whose temporal integration window is a linear superposition of an ~ 25 -ms exponential window and an ~ 200 -ms exponential window. Hence, this model can also be implemented by a single population of neurons with a complex temporal integration window. MEG may not have the spatial resolution necessary to determine whether the two temporal windows are implemented by two different neural populations or the same population. Moreover, the temporal dynamics of the aSSR recorded by MEG reflect not only the property of the MEG cortical source but also the properties of other cortical and subcortical neural populations in the same network as the MEG source. Therefore the integration-compression model does not posit any locations for the sources of temporal integration and compression.

A second type of psychoacoustical model involves a variable, task-dependent, temporal integration window (Dau et al. 1996, 1997). This model contains a “hard-wired” 20-ms temporal integration window followed by an adaptive temporal integration window whose time constant is optimized for specific tasks according to signal detection theory. The 20-ms temporal integration window models the lower limit of temporal acuity of the auditory system, whereas the adaptive optimal temporal integration window accounts for the long integration process shown in some psychoacoustical experiments (Dau et al. 1997). However, for 40-Hz aSSR, not only the fast fall-off process but also the slow build-up process are shown to be task independent (Ross and Pantev 2004; Ross et al. 2002, 2005). Hence, both the 25-ms level and the 200-ms level temporal integration windows are arguably the features of the neural circuitry in the auditory system. Of course, we acknowledge that besides the “hard-wired” temporal integration window, other types of adaptive information integration

strategies are probably adopted by high-level auditory systems and decision making systems.

A third type of psychoacoustical model is the multi-look model (Viemeister and Wakefield 1991), which suggests the auditory system takes “looks” at or “samples” (see also Poeppel 2003) from input acoustic stimuli at a fairly high rate (~ 3 ms) and selectively processes some of the looks. The advantage of this model is also its flexibility. It allows the auditory system to “intelligently” weight and process the looks stored in sensory memory. There are two possible relationships between the integration-compression model and the multi-looks model, depending on the nature of 40-Hz aSSR. The first possibility is the integration-compression model is a strategy to integrate different looks. The second possibility is each cycle of 40-Hz aSSR represents one or a chunk of “looks,” because the amplitude and phase of each cycle of 40-Hz aSSR encodes information about the acoustic stimulus. For this possibility, the integration-compression model describes the basic physical property of the auditory “sampling” system. Which possibility offers a better description depends on whether the 40-Hz aSSR is more closely related to perceptual results or to acoustic information encoding. One possible method to evaluate the functional role of 40-Hz aSSR would be to examine its dynamics after a 6-ms gap in the stimulus modulation, for trials in which the gap is not perceptible. If the amplitude and phase of the aSSR does not change after the temporal gap, it may be closer to perceptual results and the integration-compression model would need to be modified to accommodate the new “looks” integration strategy. In contrast, if the amplitude and phase of 40-Hz aSSR change because of the gap (Ross and Pantev 2004), regardless of the percept, it may be closer to the acoustic representation and can be viewed as a series of looks taken by the auditory system.

Summarizing these three types of temporal integration models, all of them assume some kind of hard-wired temporal integration window or sampling period, which reflects the basic physical property of human auditory system. The output of the hard-wired temporal integration or sampling system then interfaces with other systems, e.g., decision-making systems, speech analysis systems, and memory systems, in a task-dependent manner. The temporal dynamics of 40-Hz aSSR, as is shown in the integration-compression model, reflects the property of the hard-wired integration/sampling system. However, it is unclear whether it is only the output of the hard-wired system or is the output of a higher level system. The neural source of 40-Hz aSSR has been localized to core auditory cortex (Draganova et al. 2002; Ross et al. 2002, 2003). However, because of the limited spatial resolution of MEG, it is impossible to distinguish between the possibilities of a single source and of several nearby neural populations. However, properties of the 40-Hz aSSR are significantly different from those of neural activities recorded from mammalian primary auditory cortex. First, although the aSSR is strongest near 40 Hz, few neurons in mammalian primary cortex are tuned to 40-Hz temporal modulations (Eggermont 2002; Gaese and Ostwald 1995; Liang et al. 2002). Moreover, evoked local field potential responses in human primary auditory cortex generally have a low-pass shape with cut-off frequency < 20 Hz (Liegeois-Chauvel et al. 2004). Second, phase-locked neural activities recorded from ferret primary auditory cortex adapt out ~ 100 ms after the onset of a stationary 40-Hz temporal modulation (Elhilali

et al. 2004). However, the strong 40-Hz aSSR measured by MEG or EEG fully builds up > 200 ms after the onset of stationary temporal modulations and does not adapt even 10 min after the stimulus onset (Picton et al. 1987). More experimental work is needed to determine the function role and anatomical source of the 40-Hz aSSR.

The integration-compression model describes basic temporal properties of 40-Hz aSSR. However, it must still be refined to explain more complex experiments. For example, Ross et al. (2005) found that the amplitude of 40-Hz aSSR is temporarily reduced by a short noise burst in the continuous 40-Hz AM stimulus. This change cannot be predicted from this model. However, the time course of the amplitude fluctuation of 40-Hz aSSR induced by the short noise burst is extremely similar to that induced by a gap in the stimulus modulation. Hence, if the acoustic gating function is modified to include all acoustic changes that can elicit an amplitude fluctuation in 40-Hz aSSR, the integration-compression model may explain the result found by Ross et al. (2005).

In sum, this experiment showed the fast FM and slow AM features of the same carrier are both represented by precise temporal coding in human auditory cortex, reflected by the aSSR at f_{FM} and f_{AM} . The information of the slow AM is encoded in both the aSSR at f_{AM} and the temporal dynamics of the aSSR at f_{FM} . The temporal dynamics of the amplitude of the aSSR found in this study and previous studies (Poulsen et al. 2007; Ross and Pantev 2004; Ross et al. 2002) are successfully explained by a unified model that contains two temporal integration constants, indicating dual temporal scales exist in the auditory processing of temporal modulations.

ACKNOWLEDGMENTS

We thank D. Poeppel and M. Howard for insightful comments and J. Walker for excellent technical support.

GRANTS

This research was supported by National Institute of Deafness and Other Communication Disorders Grant R01-DC-008342.

REFERENCES

- Bendor D, Wang X.** Differential neural coding of acoustic flutter within primate auditory cortex. *Nat Neurosci* 10: 763–771, 2007.
- Boemio A, Fromm S, Braun A, Poeppel D.** Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nat Neurosci* 8: 389–395, 2005.
- Chi T, Gao Y, Guyton MC, Ru P, Shamma S.** Spectro-temporal modulation transfer functions and speech intelligibility. *J Acoust Soc Am* 106: 2719–2732, 1999.
- Dau T, Kollmeier B, Kohlrausch A.** Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration. *J Acoust Soc Am* 102: 2906–2919, 1997.
- Dau T, Puschel D, Kohlrausch A.** A quantitative model of the “effective” signal processing in the auditory system. I. Model structure. *J Acoust Soc Am* 99: 3615–3622, 1996.
- de Boer E.** Auditory time constants: a paradox? In: *Time Resolution in Auditory Systems*, edited by Michelsen A. Berlin: Springer, 1985, p. 141–158.
- de Cheveigné A, Simon JZ.** Denoising based on time-shift PCA. *J Neurosci Methods* 165: 297–305, 2007.
- de Cheveigné A, Simon JZ.** Denoising based on spatial filtering. *J Neurosci Methods* 171: 331–339, 2008a.
- de Cheveigné A, Simon JZ.** Sensor noise suppression. *J Neurosci Methods* 168: 195–202, 2008b.
- Dimitrijevic A, John MS, van Roon P, Picton TW.** Human auditory steady-state responses to tones independently modulated in both frequency and amplitude. *Ear Hear* 22: 100–111, 2001.

- Draganova R, Ross B, Borgmann C, Pantev C.** Auditory cortical response patterns to multiple rhythms of AM sound. *Ear Hear* 23: 254–265, 2002.
- Drullman R, Festen JM, Plomp R.** Effect of temporal envelope smearing on speech reception. *J Acoust Soc Am* 95: 1053–1064, 1994.
- Eggermont JJ.** Temporal modulation transfer functions in cat primary auditory cortex: separating stimulus effects from neural mechanisms. *J Neurophysiol* 87: 305–321, 2002.
- Elhilali M, Fritz JB, Klein DJ, Simon JZ, Shamma SA.** Dynamics of precise spike timing in primary auditory cortex. *J Neurosci* 24: 1159–1172, 2004.
- Ewert SD, Verhey JL, Dau T.** Spectro-temporal processing in the envelope-frequency domain. *J Acoust Soc Am* 112: 2921–2931, 2002.
- Fisher NI.** *Statistical Analysis of Circular Data*. New York: Cambridge, 1993.
- Friesen LM, Shannon RV, Baskent D, Wang X.** Speech recognition in noise as a function of the number of spectral channels: comparison of acoustic hearing and cochlear implants. *J Acoust Soc Am* 110: 1150–1163, 2001.
- Fujiki N, Jousmaki V, Hari R.** Neuromagnetic responses to frequency-tagged sounds: a new method to follow inputs from each ear to the human auditory cortex during binaural hearing. *J Neurosci* 22: 205RC, 2002.
- Fullgrabe C, Lorenzi C.** Perception of the envelope-beat frequency of inharmonic complex temporal envelopes. *J Acoust Soc Am* 118: 3757–3765, 2005.
- Gaese BH, Ostwald J.** Temporal coding of amplitude and frequency modulation in the rat auditory cortex. *Eur J Neurosci* 7: 438–450, 1995.
- Giraud A-L, Kleinschmidt A, Poeppel D, Lund TE, Frackowiak RSJ, Laufs H.** Endogenous cortical rhythms determine cerebral specialization for speech perception and production. *Neuron* 56: 1127–1134, 2007.
- Giraud A-L, Lorenzi C, Ashburner J, Wable J, Johnsrude I, Frackowiak R, Kleinschmidt A.** Representation of the temporal envelope of sounds in the human brain. *J Neurophysiol* 84: 1588–1598, 2000.
- Greenberg S, Carvey H, Hitchcock L, Chang S.** Temporal properties of spontaneous speech—a syllable-centric perspective. *J Phon* 31: 465–485, 2003.
- Hart HC, Palmer AR, Hall DA.** Amplitude and frequency-modulated stimuli activate common regions of human auditory cortex. *Cereb Cortex* 13: 773–781, 2003.
- John MS, Dimitrijevic A, Roon PV, Picton TW.** Multiple auditory steady-state responses to AM and FM stimuli. *Audiol Neurootol* 6: 12–27, 2001.
- John MS, Picton TW.** MASTER: a Windows program for recording multiple auditory steady-state responses. *Comput Methods Programs Biomed* 61: 125–150, 2000.
- Joris PX, Schreiner CE, Rees A.** Neural processing of amplitude-modulated sounds. *Physiol Rev* 84: 541–577, 2004.
- Lakatos P, Shah AS, Knuth KH, Ulbert I, Karmos G, Schroeder CE.** An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. *J Neurophysiol* 94: 1904–1911, 2005.
- Liang L, Lu T, 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, Chauvel P.** Temporal envelope processing in the human left and right auditory cortices. *Cereb Cortex* 14: 731–740, 2004.
- Lorenzi C, Gilbert G, Carn H, Garnier S, Moore BCJ.** Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. *Proc Natl Acad Sci USA* 103: 18866–18869, 2006.
- Lorenzi C, Simpson MIG, Millman RE, Griffiths TD, Woods WP, Rees A, Green GGR.** Second-order modulation detection thresholds for pure-tone and narrow-band noise carriers. *J Acoust Soc Am* 110: 2470–2478, 2001a.
- Lorenzi C, Soares C, Vonner T.** Second-order temporal modulation transfer functions. *J Acoust Soc Am* 110: 1030–1038, 2001b.
- Luo H, Wang Y, Poeppel D, Simon JZ.** Concurrent encoding of frequency and amplitude modulation in human auditory cortex: MEG evidence. *J Neurophysiol* 96: 2712–2723, 2006.
- Luo H, Wang Y, Poeppel D, Simon JZ.** Concurrent encoding of frequency and amplitude modulation in human auditory cortex: encoding transition. *J Neurophysiol* 98: 3473–3485, 2007.
- Ménard M, Gallégo S, Berger-Vachon C, Collet L, Thai-Vana H.** Relationship between loudness growth function and auditory steady-state response in normal-hearing subjects. *Hear Res* 235: 105–113, 2008.
- Millman RE, Prendergast G, Kitterick PT, Woods WP, Green GGR.** Spatiotemporal reconstruction of the auditory steady-state response to frequency modulation using magnetoencephalography. *NeuroImage* (August 21, 2009). doi:10.1016/j.neuroimage.2009.08.029.
- Moore BCJ.** *Introduction to the Psychology of Hearing*. Boston, MA: Academic Press, 2003.
- Moore BCJ, 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.
- Moore BCJ, Vickers DA, Baer T, Launer S.** Factors affecting the loudness of modulated sounds. *J Acoust Soc Am* 105: 2757–2772, 1999.
- Näätänen R, Picton T.** The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology* 24: 375–425, 1987.
- Picton TW, Skinner CR, Champagne SC, Kellett AJC, Maiste AC.** Potentials evoked by the sinusoidal modulation of the amplitude or frequency of a tone. *J Acoust Soc Am* 82: 165–178, 1987.
- Poeppel D.** The analysis of speech in different temporal integration windows: cerebral lateralization as ‘asymmetric sampling in time’. *Speech Commun* 41: 245–255, 2003.
- Poulsen C, Picton TW, Paus T.** Age-related changes in transient and oscillatory brain responses to auditory stimulation in healthy adults 19–45 years old. *Cereb Cortex* 17: 1454–1467, 2007.
- Remez RE, Rubin PE, Pisoni DB, Carrell TD.** Speech perception without traditional speech cues. *Science* 212: 947–950, 1981.
- Rosen S.** Temporal information in speech: acoustic, auditory and linguistic aspects. *Philos Trans R Soc Lond B Biol Sci* 336: 367–373, 1992.
- Ross B, Borgmann C, Draganova R, Roberts LE, 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, Draganova R, Picton TW, Pantev C.** Frequency specificity of 40-Hz auditory steady-state responses. *Hear Res* 186: 57–68, 2003.
- Ross B, Herdman AT, Pantev C.** Stimulus induced desynchronization of human auditory 40-Hz steady-state responses. *J Neurophysiol* 94: 4082–4093, 2005.
- Ross B, Pantev C.** Auditory steady-state responses reveal amplitude modulation gap detection thresholds. *J Acoust Soc Am* 115: 2193–2206, 2004.
- Ross B, Picton TW, Pantev C.** Temporal integration in the human auditory cortex as represented by the development of the steady-state magnetic field. *Hear Res* 165: 68–84, 2002.
- Saberi K, Hafter ER.** A common neural code for frequency- and amplitude-modulated sounds. *Nature* 374: 537–539, 1995.
- Shamma S.** Analysis of speech dynamics in the auditory system. In: *Dynamics of Speech Production and Perception: Life and Behavioural Sciences*, edited by Divenyi P, Greenberg S, Meyer G. Washington, DC: IOS Press, 2006, p. 335–342.
- Shannon RV, Zeng F-G, Kamath V, Wygonski J, Ekelid M.** Speech recognition with primarily temporal cues. *Science* 270: 303–304, 1995.
- Viemeister NF.** Temporal modulation transfer functions based upon modulation thresholds. *J Acoust Soc Am* 66: 1364–1380, 1979.
- Viemeister NF, Wakefield GH.** Temporal integration and multiple looks. *J Acoust Soc Am* 90: 858–865, 1991.
- Wang X, Lu T, Liang L.** Cortical processing of temporal modulations. *Speech Commun* 41: 107–121, 2003.
- Zeng F-G, Nie K, Stickney GS, Kong Y-Y, Vongphoe M, Bhargava A, Wei C, Cao K.** Speech recognition with amplitude and frequency modulations. *Proc Natl Acad Sci USA* 102: 2293–2298, 2005.