

Human Neuromagnetic Steady-State Responses to Amplitude-Modulated Tones, Speech, and Music

Satu Lamminmäki,¹ Lauri Parkkonen,¹ and Riitta Hari^{1,2}

Objectives: Auditory steady-state responses that can be elicited by various periodic sounds inform about subcortical and early cortical auditory processing. Steady-state responses to amplitude-modulated pure tones have been used to scrutinize binaural interaction by frequency-tagging the two ears' inputs at different frequencies. Unlike pure tones, speech and music are physically very complex, as they include many frequency components, pauses, and large temporal variations. To examine the utility of magnetoencephalographic (MEG) steady-state fields (SSFs) in the study of early cortical processing of complex natural sounds, the authors tested the extent to which amplitude-modulated speech and music can elicit reliable SSFs.

Design: MEG responses were recorded to 90-s-long binaural tones, speech, and music, amplitude-modulated at 41.1 Hz at four different depths (25, 50, 75, and 100%). The subjects were 11 healthy, normal-hearing adults. MEG signals were averaged in phase with the modulation frequency, and the sources of the resulting SSFs were modeled by current dipoles. After the MEG recording, intelligibility of the speech, musical quality of the music stimuli, naturalness of music and speech stimuli, and the perceived deterioration caused by the modulation were evaluated on visual analog scales.

Results: The perceived quality of the stimuli decreased as a function of increasing modulation depth, more strongly for music than speech; yet, all subjects considered the speech intelligible even at the 100% modulation. SSFs were the strongest to tones and the weakest to speech stimuli; the amplitudes increased with increasing modulation depth for all stimuli. SSFs to tones were reliably detectable at all modulation depths (in all subjects in the right hemisphere, in 9 subjects in the left hemisphere) and to music stimuli at 50 to 100% depths, whereas speech usually elicited clear SSFs only at 100% depth.

The hemispheric balance of SSFs was toward the right hemisphere for tones and speech, whereas SSFs to music showed no lateralization. In addition, the right lateralization of SSFs to the speech stimuli decreased with decreasing modulation depth.

Conclusions: The results showed that SSFs can be reliably measured to amplitude-modulated natural sounds, with slightly different hemispheric lateralization for different carrier sounds. With speech stimuli, modulation at 100% depth is required, whereas for music the 75% or even 50% modulation depths provide a reasonable compromise between the signal-to-noise ratio of SSFs and sound quality or perceptual requirements. SSF recordings thus seem feasible for assessing the early cortical processing of natural sounds.

Key words: Amplitude modulation, Auditory, Frequency tagging, Magnetoencephalography, Natural stimuli.

(*Ear & Hearing* 2014;35:461–467)

¹Brain Research Unit, O.V. Lounasmaa Laboratory and MEG Core, and ²Advanced Magnetic Imaging Centre, Aalto Neuroimaging, School of Science, Aalto University, Espoo, Finland.

Copyright © 2014 by Lippincott Williams & Wilkins. This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 3.0 License, where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially.

INTRODUCTION

Auditory steady-state responses can be elicited by periodically repeated sounds, such as trains of tone bursts or clicks, or amplitude- or frequency-modulated tones (Galambos et al. 1981; Kuwada et al. 1986; Mäkelä & Hari 1987; Picton et al. 1987). Steady-state potentials (SSPs) measured by scalp electroencephalography (EEG), have been used to study both subcortical and cortical auditory processing whereas magnetoencephalographic (MEG) steady-state fields (SSFs), generated in the supratemporal auditory cortex (Mäkelä & Hari 1987; Hari et al. 1989), mainly reflect activity in the Heschl gyrus, in the primary auditory cortex (Gutschalk et al. 1999), and are therefore appropriate for studying early cortical auditory processing.

SSPs/SSFs to pure-tone stimuli (for a review, see Picton et al. 2003) reflect properties of the carrier tones: for example, cortical sources of SSFs are tonotopically organized (Pantev et al. 1996; Ross et al. 2000), and steady-state responses to stimuli with different carrier frequencies elicit responses of different amplitudes, reflecting differences in source locations or source strengths (Galambos et al. 1981; Pantev et al. 1996; Ross et al. 2000; Ross 2008). SSPs to amplitude- and/or frequency-modulated tones are routinely used in clinical work to evaluate frequency-specific hearing thresholds in an objective way (see e.g., Kuwada et al. 1986; Dimitrijevic et al. 2004; John et al. 2004), and SSFs to amplitude-modulated tones have been applied to scrutinize binaural interaction by means of “frequency-tagging” where left- and right-ear sounds are amplitude-modulated at slightly different frequencies (Fujiki et al. 2002; Kaneko et al. 2003; Lamminmäki et al. 2012a, 2012b).

However, in our daily lives we typically encounter sounds that are physically much more complex than pure tones. For example, speech and music are acoustically discontinuous and dynamic with many frequency components. Although much of the brain processing of, for example, speech relies on top-down predictions, proper perception of speech and music requires that the fine spectrotemporal features of these sounds are analyzed precisely and rapidly in the auditory system. SSFs, reflecting early sound-evoked cortical activity, might thus be useful in following cortical processing of natural sounds in everyday-life-like experimental setups.

Recently, SSPs have been recorded to amplitude-modulated (AM) speech-like stimuli (Keitel et al. 2011; Poelmans et al. 2012), but it is still unclear to which extent it is possible to record reliable steady-state responses to natural sounds while preserving the intelligibility of speech or the quality of music despite the distorting amplitude modulation.

To explore whether natural, physically complex sounds can be used as carrier sounds to elicit reliable SSFs, we recorded MEG responses to AM speech, music, and tones. We also tested the dependence of the SSFs, and the intelligibility and musical quality of the stimuli on the AM depth.

SUBJECTS AND METHODS

Subjects

We studied 11 healthy, normal-hearing volunteers (4 females, 7 males; 20 to 39 years of age; mean age 25.8 years). All subjects were right-handed according to the Edinburgh Handedness Inventory questionnaire (Oldfield 1971); mean laterality quotient was 90 (range 71 to 100).

The MEG recordings and behavioral tests had a prior approval by the Ethics Committee of Helsinki Uusimaa Hospital District, and the subjects signed a written consent form after a thorough explanation of the experiment.

Stimuli

Carrier sounds included natural, nonpreprocessed speech (Plato's *The Republic*, Naxos AudioBooks), and violin music (G. Tartini's Sonata for violin in G minor *Le trille de diable*), and a continuous 1-kHz tone. Each sound was amplitude modulated at 41.1 Hz at four modulation depths (100, 75, 50, and 25%) by a custom program running on Matlab (MathWorks, Natick, MA). For example, the 1-kHz tone was modulated with 75% depth according to formula:

$$\text{tone}_{75}(t) = \sin(2\pi \times 1000 \times t) \times (1 + \sin(2\pi \times 41.1 \times t) \times 0.75)$$

where t is time. Modulation frequencies near 40 Hz are known to elicit strong SSRs (Galambos et al. 1981; Hari et al. 1989), and 41.1 Hz was selected on the basis of an earlier frequency-tagging study to minimize line-frequency interference (Kaneko et al. 2003).

Figure 1 shows short epochs of the original acoustic speech signal (top trace; AM 0%) and the same signals amplitude-modulated at 25, 50, 75, and 100%.

The presentation order of the 12 stimuli (tone, music, and speech, each with 4 modulation depths) was randomized individually, and the stimulus sequence was repeated once in the reverse order. In the following, subscripts 25, 50, 75, and 100 after the stimulus names are used to refer to the modulation depths (e.g., tone_{75} , music_{25} , speech_{100}). The 90-s stimuli were separated by 5 s, and the pause between the two stimulus runs was 5 min. To determine the noise level, we recorded 2 min of MEG without stimulation. The whole experiment lasted for about 45 min.

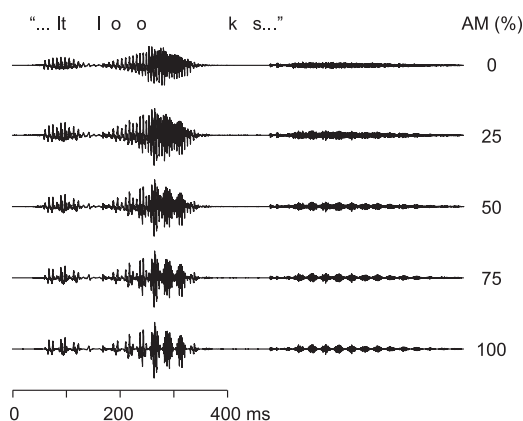


Fig. 1. Speech stimuli. The original speech waveform (top trace), amplitude modulated at 0%; and the same sound amplitude modulated at 25, 50, 75, and 100%. AM indicates amplitude-modulated.

Experimental Setup

As the result of the amplitude modulation, the root-mean-square (RMS) level of the stimulus was the highest at the 25% modulation and the lowest at the 100% modulation (as the modulated signal attained much lower values at 100% than 25% modulation). To rule out possible effects of the resulting slight intensity differences between the stimuli, we performed a control experiment in 2 subjects by recording SSFs also to RMS-equalized sounds.

The subjects were instructed to listen to the binaurally presented sounds and to keep their eyes open. The hearing thresholds were tested separately for both ears, and the stimulation was set individually to a comfortable listening level (50 to 65 dB above the sensation threshold).

MEG Recording

The brain's magnetic signals were recorded with a whole-scalp neuromagnetometer (Elekta-Neuromag Vectorview; Elekta Oy, Helsinki, Finland), which comprises 204 first-order planar gradiometers and 102 magnetometers. The signals were band-pass filtered to 0.1 to 200 Hz and digitized at 600 Hz. Altogether circa 40 min of data were gathered. Vertical and horizontal electro-oculograms were recorded during the measurement.

Behavioral Test

Eight of the 11 subjects participated in the behavioral test after the MEG recording. Sixteen 15-s speech fragments (4 of each modulation depth) were presented binaurally, and the subjects were asked to evaluate the intelligibility of the speech on visual analog scale from 0 (noise, unable to understand) to 100 (normal, nonmodulated speech); examples of the extremes were presented before the test. The musical quality of the music stimuli was tested similarly (0 = noise, not music; 100 = the original nonmodulated violin music). Finally, the subjects compared speech_{100} and music_{100} with the corresponding nonmodulated sounds and evaluated them in terms of naturalness (0 = very unnatural, artificial; 100 = natural, equal to the nonmodulated sound) and in terms of subjective deterioration caused by the modulation (0 = very deteriorated; 100 = no deterioration).

Time-Domain Analysis of MEG Data

To unravel cortical SSFs, about 1300 MEG epochs (146 ms each, corresponding to 6 periods of the 41.1-Hz modulation frequency; overlapping by 4 cycles) were averaged in phase with the modulating signal. The averaged signals were band-pass filtered to 32 to 48 Hz. All MEG epochs coinciding with eye blinks (electro-oculograms > 150 μV) were omitted before averaging. The measured noise data (i.e., MEG signals during silence) were averaged and filtered similarly as the responses.

Neuronal sources of SSFs to tone_{100} were modeled using a well-established least-squares-fit procedure (for a review, see Hämäläinen et al. 1993) with two equivalent current dipoles (ECDs), one in each hemisphere. ECDs explaining over 85% of the signal variance over the auditory cortex (28 planar gradiometer over each temporal lobe) were accepted for further analysis. Due to insufficient signal-to-noise ratio, accurate spatial parameters of the ECDs could not be fitted separately for responses to all modulation depths and sound types, especially for responses to speech and music at the most shallow amplitude modulations. Therefore, we considered two current dipoles fitted for tone_{100} , fixed in location and direction, to be

the most reliable approach to explain the responses to all stimuli. The fidelity of the sources was verified in two representative subjects for whom ECDs were fitted separately for tone₁₀₀, tone₇₅, music₁₀₀, music₇₅, speech₁₀₀, and speech₇₅: for all those stimuli, the source locations were similar in both hemispheres. In most conditions, ECDs fitted to tone₁₀₀ explained the field patterns as well as or even slightly better than the ECDs fitted to responses with weaker signal-to-noise ratio. The dipole locations in proportion to anatomical structures were identified in 7 subjects on the basis of magnetic resonance images, and the source strengths were measured as peak-to-peak values from source waveforms.

Laterality indices (LIs) of the individual source strengths were calculated according to the formula $LI = (R - L)/(R + L)$, where R refers to the source strength in the right hemisphere (RH) and L in the left hemisphere (LH).

Frequency-Domain Analysis of MEG Signals

For spectral analysis, the raw MEG signals from the 90-s stimulation epochs were down-sampled by a factor of three. For each 204 planar gradiometer channel, half-overlapping Hanning-windowed 4096-point Fast Fourier Transforms (spectral bin width 0.01 Hz) were computed separately over the whole 90-s stimulation epoch, resulting in seven spectra (one for each windowed segment). The magnitudes of the resulting seven spectra were averaged. In each hemisphere, the strengths of the spectral peaks at 41.1 Hz were quantified from an individually selected temporal channel showing the strongest 41.1-Hz activity in the time-domain average.

Statistical Analysis

Statistical significances were assessed using Student paired *t*-tests and repeated-measures analyses of variance (ANOVA). In the ANOVA of source strengths and spectral peaks, a $2 \times 3 \times 4$ (hemisphere \times stimulus type \times modulation depth) design was used whereas LIs were analyzed with a 3×4 (Stimulus Type \times Modulation Depth) design. When the assumption of sphericity was violated, Greenhouse–Geisser-corrected values were applied in ANOVA. In cases of multiple *t*-tests, Bonferroni corrections were used; the significance level of each individual *t*-test was set to $0.05/n$ where *n* was the number of comparisons.

RESULTS

Behavioral Results

The intelligibility of speech decreased with the increasing modulation depth (see the upper panel of Fig. 3): at 100% depth, intelligibility was clearly impeded whereas the 25% modulation depth had only a minor effect (mean \pm standard error of the mean [SEM] visual analog scale, 54.4 ± 8.1 versus 94.1 ± 2.6 , $p = 0.001$; value 0 corresponding to meaningless noise and value 100 to nonmodulated speech). However, all subjects—although nonnative English speakers—were able to understand the contents of all speech stimuli, even speech₁₀₀.

For music, increasing modulation depth degraded the musical quality, which was significantly lower for music₁₀₀ than for music₂₅ (Fig. 3, upper part, 55.6 ± 11.4 versus 84.0 ± 5.1 , $p < 0.02$; value 0 corresponding to noise and 100 to nonmodulated music).

The naturalness scores did not differ for speech₁₀₀ versus music₁₀₀ (43.8 ± 7.4 versus 48.1 ± 6.5). However, the modulation

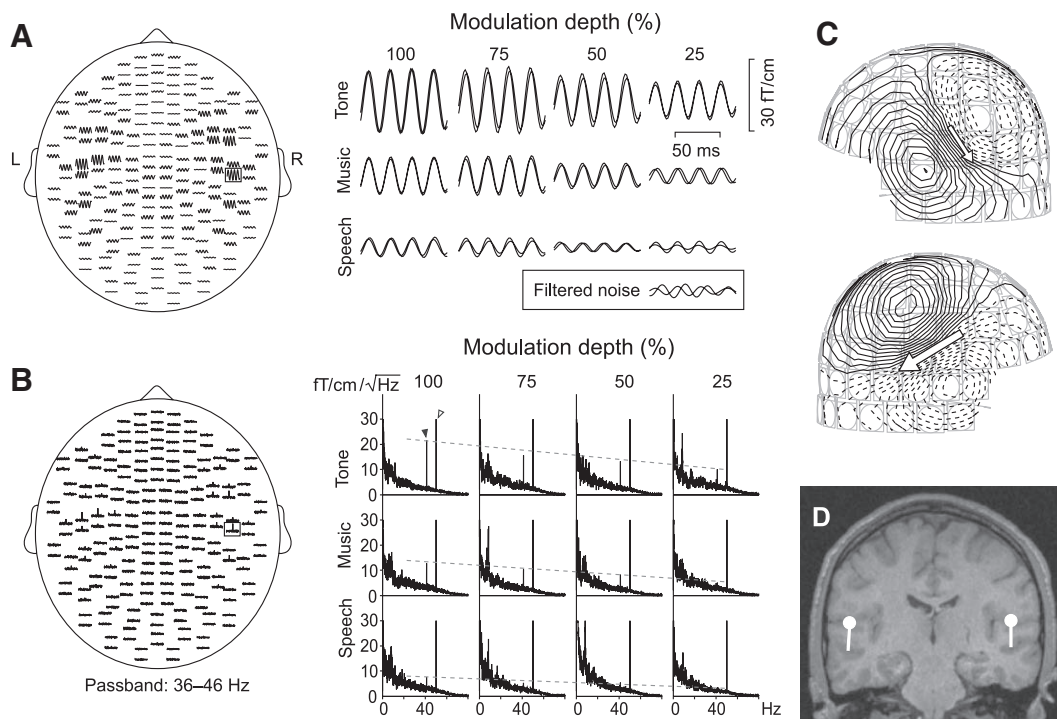


Fig. 2. Steady-state responses in a single subject. A, left, SSFs on planar gradiometer channels to a binaural 1-kHz tone, amplitude-modulated at 41.1 Hz at 100% modulation depth (tone₁₀₀). A, right, Single-channel SSFs to all 12 stimuli. The inset shows filtered noise from the same channel, measured from a subject receiving no sound stimuli. B, The corresponding frequency-domain signals. The dashed lines follow the spectral peaks at the modulation frequency of 41.1 Hz. Power line artifact is seen at 50 Hz, and spontaneous rhythmic brain activity below 25 Hz. C, Magnetic field pattern of SSFs to tone₁₀₀; isocontour lines are separated by 3 fT. D, Two equivalent current dipoles best explaining the field pattern of (C), displayed on subject's own magnetic resonance image.

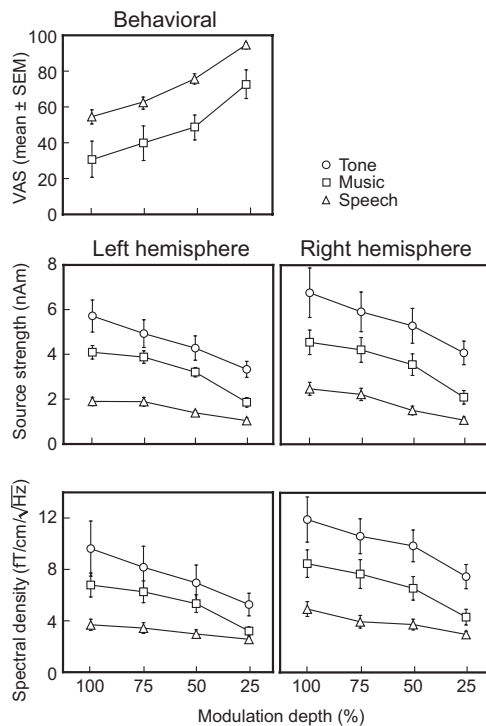


Fig. 3. Behavioral ratings and response strengths for all stimuli. Mean \pm SEM visual analog scale scores (top), the mean \pm SEM source strengths for steady-state fields (middle) and the mean \pm SEM spectral amplitudes (bottom). SEM indicates standard error of the mean.

deteriorated music much more than speech (music₁₀₀ 27.8 \pm 8.5 versus speech₁₀₀ 48.4 \pm 7.5, $p < 0.003$). During debriefing, 5 of the 8 tested subjects reported spontaneously that although both music₁₀₀ and speech₁₀₀ sounded unnatural (according to a few subjects like a crackly radio), the modulation did not prevent following the content of speech; instead, the concentration on the melody of the music was clearly disrupted.

Individual Responses

The stimuli elicited clear SSFs and spectral peaks over both auditory cortices. Figure 2 displays responses of one subject: the averaged band-pass filtered (32 to 48 Hz) SSFs in time-domain (Fig. 2A) and the frequency-domain signals (Fig. 2B). These spatial distributions agree with our previous studies (Fujiki et al. 2002; Kaneko et al. 2003; Lamminmäki et al. 2012b).

SSFs to all 12 applied stimuli diminished as the modulation depth decreased from 100 to 25%, as is evident from the single-channel responses in Figure 2A (right). The tones which were physically simplest and least variable of our stimuli elicited the strongest responses at each modulation depth whereas the speech sounds, which were physically most complex and variable of our stimuli, elicited the weakest responses.

The corresponding single-channel amplitude spectra (Fig. 2B, right) demonstrated similar effects as the time-domain SSFs: the spectral peaks at the modulation frequency (41.1 Hz) decreased when the modulation depth decreased or the stimulus complexity increased (from tone to music to speech). Peaks at frequencies lower than 25 Hz derive from spontaneous brain activity, and the power line artifact appears at 50 Hz.

Whereas tone₁₀₀ produced strong and clear responses, the responses to speech₂₅ were not discernible from filtered

measurement noise (the inset in the lower right corner in Fig. 2A), and no distinct spectral peak was visible at 41.1 Hz. However, unlike filtered noise, these weak responses to speech₂₅ were still approximately in phase during the two recordings and their field patterns were roughly dipolar.

Fig. 2C displays the clearly dipolar magnetic field patterns for SSFs to tone₁₀₀. The sources were adequately modeled with two ECDs (marked by arrows), one in each hemisphere (Fig. 2D).

Source Locations and Strengths

In all the 7 subjects whose anatomical magnetic resonance images were available, the current dipoles for SSFs were verified to reside in the supratemporal auditory cortices, in line with earlier reports on auditory SSFs (Mäkelä & Hari 1987; Hari et al. 1989; Gutschalk et al. 1999). The sources of tone₁₀₀ were on average 6 mm ($p < 0.009$, Bonferroni-corrected paired t -test) more anterior in the RH than the LH; in inferior–superior or medial–lateral directions no differences between the hemispheres were found.

Figure 3 (middle) shows the mean \pm SEM source strengths of SSFs to all stimuli, separately in both hemispheres. The source strengths depended both on the stimulus type ($F(2,20) = 41.87$, $p < 0.0005$, partial $\eta^2 = 0.81$) and modulation depth ($F(3,30) = 46.07$, $p < 0.0005$, partial $\eta^2 = 0.82$) similarly to the individual responses exemplified in Figure 2. The hemisphere had no main effect on the source strengths ($F(1,10) = 2.05$, $p = 0.18$).

In LH, the source strengths were 53 \pm 3% (mean \pm SEM) weaker to speech₁₀₀ than to music₁₀₀, and 24 \pm 5% weaker to music₁₀₀ than to tone₁₀₀. In the RH, the corresponding percentages were 45 \pm 3% and 28 \pm 4%, respectively.

Spectra

According to visual examination of individual data, clear spectral peaks above the noise level were elicited by all tone stimuli (in the RH of all 11 subjects, in the LH of 9 subjects), as well as by the music stimuli modulated at 50 to 100% depths (in the RH of all 11 subjects, in the LH of 10 subjects at 75 to 100% depths and of 9 subjects at 50% depth). The speech stimuli elicited clear spectral peaks only with the strongest modulation: at 75 and 100% depths in both hemispheres of 5 subjects, only at 100% depth in both hemispheres of 4 subjects as well as in the RH of 1 subject; in 1 subject, no clear spectral peaks were found. At 25% modulation depth, speech elicited clear spectral peaks in both hemispheres of 1 subject and in one hemisphere of 4 subjects.

The bottom part of Figure 3 demonstrates the mean \pm SEM amplitudes of the spectral peaks. In line with the source modeling results, the amplitudes of the spectral peaks depended both on the stimulus type ($F(2,20) = 24.12$, $p < 0.0005$, partial $\eta^2 = 0.71$) and the modulation depth ($F(3,30) = 27.28$, $p < 0.0005$, partial $\eta^2 = 0.73$): the tone stimuli produced the strongest spectral peaks and the speech stimuli the weakest, and the amplitudes decreased when the modulation depth decreased.

The spectral estimates were derived separately from the measured raw data for all conditions whereas the source strengths were always computed with the equivalent current dipole model derived for tone₁₀₀. Pearson correlation coefficients showed that the source strengths and spectral peaks were strongly correlated in each hemisphere (mean \pm SEM correlation coefficients across subjects 0.91 \pm 0.03 in the LH and 0.96 \pm 0.02 in the RH), thereby demonstrating the similarity of these two analysis approaches in quantifying the response magnitudes.

Hemispheric Lateralization

Figure 4 shows the LIs of SSF source strengths to stimuli modulated at 50 to 100%. The stimulus type (tone, music, speech) had a near-significant main effect on the LIs ($F(2,20) = 3.38, p = 0.054, \text{partial } \eta^2 = 0.25$): SSFs to tone stimuli and to speech stimuli were statistically significantly lateralized to the RH (for each carrier, $p = 0.04$ when compared with 0 level, Bonferroni corrected against multiple comparisons), whereas LIs to music stimuli did not show significant lateralization.

In addition, a significant interaction existed between the stimulus type (tone, music, speech) and the modulation depth ($F(6,60) = 3.35, p = 0.026, \text{partial } \eta^2 = 0.25$). First, with 100% modulation depth, LIs to tone and music stimuli did not differ statistically significantly from zero, whereas LIs for speech were lateralized to the RH ($p < 0.05$, Bonferroni corrected) and differed from LIs to music (0.12 ± 0.03 versus 0.04 ± 0.04 , $p = 0.004$, Bonferroni-corrected paired t -test). Second, the modulation depth had no effect on the LIs for tones or music, whereas for speech the responses were statistically significantly more RH lateralized at 100% than at 50% modulations ($p = 0.050$, Bonferroni-corrected paired t -test).

In the control experiment carried out with RMS-equalized stimuli on 2 subjects, the order of the source strengths and their lateralization were as in the main study, thereby ruling out the effect of slight intensity differences between the stimuli.

DISCUSSION

Our results demonstrate that not only AM tones but also similarly modulated music and speech stimuli, with tolerable quality and intelligibility, elicit reliable SSFs when the amplitude modulation is deep enough. The source strengths and hemispheric lateralization of SSFs depended on the carrier sounds, suggesting that SSFs can inform about the cortical processing of natural sounds.

Two-Sided Effects of Modulation Depth

With all sound types, the SSFs increased in both hemispheres with the increasing modulation depth, replicating earlier findings for pure tones (Kuwada et al. 1986; Rees et al. 1986; Ross et al. 2000; Picton et al. 2003). Series of clear on-off sounds, such as clicks, tones, or noise bursts, are known to

trigger strong steady-state responses, and at the deepest modulation, our stimuli started to resemble these efficient stimuli (although for speech and music the sound amplitude varied somewhat during the on-phase). Thus it was expected that the sounds with the deepest modulation elicited SSFs with the highest signal-to-noise ratio. It was also expected that simultaneously the speech intelligibility and the musical quality of the stimuli decrease, as an increasing part of the original stimulus is periodically dampened. In other words, the cost of the higher signal-to-noise ratio of the SSFs is the unnaturalness and discontinuity of the sound stimulus. Or on the positive side: even when speech and music were amplitude-modulated so much that they elicited reliable SSFs, they still remained intelligible.

As could be expected on the basis of the physical complexity of the carrier sounds, SSFs were weaker to music than to tones and the weakest to speech stimuli. Respectively, reliable SSFs were obtained to tones at all modulation depths from 25 to 100%, whereas depths of at least 50% were needed for music and at least 75% for speech. However according to the subject's ratings, the modulation disrupted more the quality of music than speech; speech was comprehensible even at 100% modulation, agreeing with the more important role of the amplitude envelopes than spectral cues in speech recognition (Shannon et al. 1995). To maximize SSFs (and signal-to-noise ratio) in future studies involving naturalistic sounds, the modulation should be as deep as feasible, to the extent allowed by the sound quality and perceptual requirements of the experimental setup. In studying early cortical processing of speech, the 100% modulation depth might be the most reasonable choice, whereas with music, the 75% or even 50% modulation depth might be a suitable compromise between the signal-to-noise ratio of SSFs and perceptual fidelity.

Hemispheric Lateralization of SSFs

We found that SSFs to tones and speech stimuli were lateralized to the RH, whereas SSFs to music showed no lateralization. In addition, at 100% AM, corresponding to the best signal-to-noise ratio, SSFs to tone stimulus were the most RH lateralized and differed significantly from SSFs to music. These lateralization differences between the conditions support the earlier findings that SSFs reflect properties of the carrier sounds (Galambos et al. 1981; Pantev et al. 1996; Ross et al. 2000; Ross 2008).

Regarding pure tones, our results conform to the findings of Ross et al. (2005) who showed that the SSFs to a binaural 0.5 kHz AM tone lateralize slightly toward RH (mean LI 0.219), although the hemispheric difference depends strongly on the subject (LI range from +0.6 to -0.2, in our study from +0.4 to -0.3). The RH lateralization of SSFs to natural speech might seem contradictory to a multitude of neuroimaging studies that have demonstrated a more important role of LH than RH in language processing (for a review, see e.g., Friederici & Alter 2004; Friederici 2011). However, speech, as all sounds, is processed bilaterally, and RH contributes substantially to processing of prosodic information and therefore to the comprehension of speech (for a review, see Jung-Beeman 2005; Zatorre & Gandour 2008; Friederici 2011). In addition, brain processing of complex sounds is hierarchical (for a review, see e.g., Griffiths 2001), and the well-known LH preference in speech processing is related to higher-order processing steps whereas steady-state

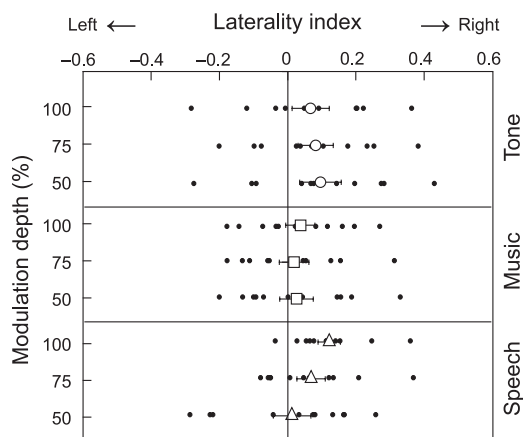


Fig. 4. Laterality indices for tone, music, and speech stimuli. Individual indices for stimuli modulated at 50 to 100% are marked with dots, and the mean (\pm standard error of the mean) values with larger symbols.

responses reflect lower-level processing in the primary auditory cortex and its vicinity (Mäkelä & Hari 1987; Hari et al. 1989; Gutschalk et al. 1999). Therefore SSFs may provide complementary information about early cortical processing of natural sounds, including speech.

In the present study, SSRs to AM music showed no hemispheric lateralization. Listening to music typically evokes stronger and more widespread activation in the RH than in the LH (see e.g., Zatorre et al. 2002; Tervaniemi & Hugdahl 2003; Alluri et al. 2013). However, the SSRs reflect mostly the early processing of basic acoustic features and—at least at this high modulation frequency—cannot span the musical structure of the sound, which may explain the lack of such lateralization.

For tones and music, the hemispheric lateralization did not depend on the modulation depth, whereas for speech, the clear RH dominance diminished with decreasing modulation depth. This weak, yet statistically significant, shift in lateralization may result from stronger weakening of the responses from 100 to 50% depth in the RH than in the LH (see Fig. 3). Although both hemispheres are important for processing of both spectral and temporal aspects of sounds, the RH is particularly important for processing of spectral features, melody, pitch of complex sounds, and prosodic and paralinguistic aspects of speech (Zatorre & Belin 2001; Zatorre et al. 2002; Friederici & Alter 2004).

The decreasing modulation makes the periodicity of the sound less clear, which might weaken the activation more in the RH than in the LH. However, the present study cannot distinguish whether the lateralization shift of speech—as well as the other differences we found in LIs between the three carrier sounds—reflects solely differences in early cortical processing of simple acoustic features or whether the responses also receive top-down contribution from the higher-level processing areas.

The role of the LH in processing syntactic and semantic information of speech is well established (for a review, see Friederici & Alter 2004; Lindell 2006), and the increased naturalness of the speech at shallower modulations could shift the hemispheric balance toward the LH. A similar lateralization shift from right to left, measured by EEG, has been found earlier for a stimulus continuum ranging from a strongly filtered tone-like sound to a phonetic sound (vowel; Rinne et al. 1999). After all, one possibility is that the activity weakens equally in both hemispheres and that the weak shift in lateralization results from the low signal-to-noise ratio in the LH, which hampers the detection of changes especially at shallow amplitude modulation.

CONCLUSION

Our findings that SSFs can be measured reliably to AM speech and music, which still have reasonably good sound quality, imply that SSFs to these natural sounds might be applied in both basic and clinical research of human brain function. Further recordings of steady-state responses to amplitude modulation of natural sounds and spectrally and temporally matched, similarly modulated nonspeech/nonmusic stimuli are needed to unravel the full capacity of this approach.

ACKNOWLEDGMENTS

The study was financially supported by the Academy of Finland (grant #129678 from the National Centers of Excellence Programme 2006–2011, and grants #131483 and #263800), the Sigrid Jusélius Foundation, the ERC

Advanced Grant #232946, and the aivoAALTO research project of the Aalto University.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the article.

The authors declare no other conflict of interest.

Address for correspondence: Satu Lamminmäki, Brain Research Unit, O.V. Louasmaa Laboratory, School of Science, Aalto University, P.O. Box 15100, FI-00076 AALTO, Espoo, Finland. E-mail sala@neuro.hut.fi

Received September 26, 2012; accepted December 18, 2013.

REFERENCES

- Alluri, V., Toivianen, P., Lund, T. E., et al. (2013). From Vivaldi to Beatles and back: Predicting lateralized brain responses to music. *Neuroimage*, *83*, 627–636.
- Dimitrijevic, A., John, M. S., Picton, T. W. (2004). Auditory steady-state responses and word recognition scores in normal-hearing and hearing-impaired adults. *Ear Hear*, *25*, 68–84.
- Friederici, A. D. (2011). The brain basis of language processing: From structure to function. *Physiol Rev*, *91*, 1357–1392.
- Friederici, A. D., & Alter, K. (2004). Lateralization of auditory language functions: A dynamic dual pathway model. *Brain Lang*, *89*, 267–276.
- Fujiki, N., Jousmäki, V., Hari, R. (2002). 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*, RC205.
- Galambos, R., Makeig, S., Talmachoff, P. J. (1981). A 40-Hz auditory potential recorded from the human scalp. *Proc Natl Acad Sci U S A*, *78*, 2643–2647.
- Griffiths, T. D. (2001). The neural processing of complex sounds. *Ann NY Acad Sci*, *930*, 133–142.
- Gutschalk, A., Mase, R., Roth, R., et al. (1999). Deconvolution of 40 Hz steady-state fields reveals two overlapping source activities of the human auditory cortex. *Clin Neurophysiol*, *110*, 856–868.
- Hämäläinen, M., Hari, R., Ilmoniemi, R. J., et al. (1993). Magnetoencephalography—Theory, instrumentation, and applications to noninvasive studies of the working human brain. *Rev Mod Phys*, *65*, 413–497.
- Hari, R., Hämäläinen, M., Joutsiniemi, S. L. (1989). Neuromagnetic steady-state responses to auditory stimuli. *J Acoust Soc Am*, *86*, 1033–1039.
- John, M. S., Brown, D. K., Muir, P. J., et al. (2004). Recording auditory steady-state responses in young infants. *Ear Hear*, *25*, 539–553.
- Jung-Beeman, M. (2005). Bilateral brain processes for comprehending natural language. *Trends Cogn Sci*, *9*, 512–518.
- Kaneko, K., Fujiki, N., Hari, R. (2003). Binaural interaction in the human auditory cortex revealed by neuromagnetic frequency tagging: No effect of stimulus intensity. *Hear Res*, *183*, 1–6.
- Keitel, C., Schröger, E., Saupe, K., et al. (2011). Sustained selective intermodal attention modulates processing of language-like stimuli. *Exp Brain Res*, *213*, 321–327.
- Kuwada, S., Batra, R., Maher, V. L. (1986). Scalp potentials of normal and hearing-impaired subjects in response to sinusoidally amplitude-modulated tones. *Hear Res*, *21*, 179–192.
- Lamminmäki, S., Mandel, A., Parkkonen, L., et al. (2012a). Binaural interaction and the octave illusion. *J Acoust Soc Am*, *132*, 1747–1753.
- Lamminmäki, S., Massinen, S., Nopola-Hemmi, J., et al. (2012b). Human ROBO1 regulates interaural interaction in auditory pathways. *J Neurosci*, *32*, 966–971.
- Lindell, A. K. (2006). In your right mind: Right hemisphere contributions to language processing and production. *Neuropsychol Rev*, *16*, 131–148.
- Mäkelä, J. P., & Hari, R. (1987). Evidence for cortical origin of the 40 Hz auditory evoked response in man. *Electroencephalogr Clin Neurophysiol*, *66*, 539–546.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*, 97–113.
- Pantev, C., Roberts, L. E., Elbert, T., et al. (1996). Tonotopic organization of the sources of human auditory steady-state responses. *Hear Res*, *101*, 62–74.
- Picton, T. W., John, M. S., Dimitrijevic, A., et al. (2003). Human auditory steady-state responses. *Int J Audiol*, *42*, 177–219.
- Picton, T. W., Skinner, C. R., Champagne, S. C., et al. (1987). Potentials evoked by the sinusoidal modulation of the amplitude or frequency of a tone. *J Acoust Soc Am*, *82*, 165–178.

- Poelmans, H., Luts, H., Vandermosten, M., et al. (2012). Auditory steady-state cortical responses indicate deviant phonemic-rate processing in adults with dyslexia. *Ear Hear*, *33*, 134–143.
- Rees, A., Green, G. G., Kay, R. H. (1986). Steady-state evoked responses to sinusoidally amplitude-modulated sounds recorded in man. *Hear Res*, *23*, 123–133.
- Rinne, T., Alho, K., Alku, P., et al. (1999). Analysis of speech sounds is left-hemisphere predominant at 100–150ms after sound onset. *Neuroreport*, *10*, 1113–1117.
- Ross, B. (2008). A novel type of auditory responses: Temporal dynamics of 40-Hz steady-state responses induced by changes in sound localization. *J Neurophysiol*, *100*, 1265–1277.
- Ross, B., Borgmann, C., Draganova, R., et al. (2000). A high-precision magnetoencephalographic study of human auditory steady-state responses to amplitude-modulated tones. *J Acoust Soc Am*, *108*, 679–691.
- Ross, B., Herdman, A. T., Pantev, C. (2005). Right hemispheric laterality of human 40 Hz auditory steady-state responses. *Cereb Cortex*, *15*, 2029–2039.
- Shannon, R. V., Zeng, F. G., Kamath, V., et al. (1995). Speech recognition with primarily temporal cues. *Science*, *270*, 303–304.
- Tervaniemi, M., & Hugdahl, K. (2003). Lateralization of auditory-cortex functions. *Brain Res Brain Res Rev*, *43*, 231–246.
- Zatorre, R. J., & Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cereb Cortex*, *11*, 946–953.
- Zatorre, R. J., & Gandour, J. T. (2008). Neural specializations for speech and pitch: Moving beyond the dichotomies. *Philos Trans R Soc Lond B Biol Sci*, *363*, 1087–1104.
- Zatorre, R. J., Belin, P., Penhune, V. B. (2002). Structure and function of auditory cortex: Music and speech. *Trends Cogn Sci*, *6*, 37–46.