Representation of Auditory-Filter Phase Characteristics in the Cortex of Human Listeners

André Rupp,1 Norman Sieroka,2 Alexander Gutschalk,1 and Torsten Dau3

1Section of Biomagnetism, Department of Neurology, University of Heidelberg, Heidelberg, Germany; 2Chair for Philosophy, Eidgenössische Technische Hochschule Zürich, Zurich, Switzerland; and 3Centre for Applied Hearing Research, Department of Electrical Engineering, Technical University of Denmark, Lyngby, Denmark

Submitted 10 July 2007; accepted in final form 6 January 2008

Rupp A, Sieroka N, Gutschalk A, Dau T. Representation of auditory-filter phase characteristics in the cortex of human listeners. J Neurophysiol 99: 1152–1162, 2008. First published January 9, 2008; doi:10.1152/jn.00778.2007. Harmonic tone complexes with component phases, adjusted using a variant of a method proposed by Schroeder, can produce pure-tone masked thresholds differing by >20 dB. This phenomenon has been qualitatively explained by the phase characteristics of the auditory filters on the basilar membrane, which differently affect the flat envelopes of the Schroeder-phase maskers. We examined the influence of auditory-filter phase characteristics on the neural representation in the auditory cortex by investigating cortical auditory evoked fields (AEFs). We found that the P1m component exhibited larger amplitudes when a long-duration tone was presented in a repeating linearly downward sweeping (Schroeder positive, or $m_+$) masker than in a repeating linearly upward sweeping (Schroeder negative, or $m_-$) masker. We also examined the neural representation of short-duration tone pulses presented at different temporal positions within a single period of three maskers differing in their component phases ($m_+$, $m_-$, and sine phase $m_0$). The P1m amplitude varied with the position of the tone pulse in the masker and depended strongly on the masker waveform. The neuromagnetic results in all cases were consistent with the perceptual data obtained with the same stimuli and with results from simulations of neural activity at the output of cochlear preprocessing. These findings demonstrate that phase effects in peripheral auditory processing are accurately reflected up to the level of the auditory cortex.

INTRODUCTION

One of the earliest and most fundamental stages of auditory processing is the frequency analysis that takes place in the cochlea. This stage has often been modeled by a bank of overlapping filters, and much effort has gone into characterizing the magnitude response of the “auditory filters” (e.g., Fletcher 1940; Glasberg and Moore 1990; Patterson 1976; Plomp 1964; Zwicker et al. 1957). Conversely, filter phase responses received much less attention, partly due to the popular belief that the ear is essentially “phase deaf” (Helmholtz 1954) and the observation that many perceptual properties can be explained by analyzing the power spectrum of a sound—i.e., by discarding the phase information. Although a number of studies have since shown that the ear is sensitive to changes in stimulus phase, both within an auditory filter (Goldstein 1967; Mathes and Miller 1947; Zwicker 1952) and, to a lesser extent, across filters (Patterson 1987), they have mostly assumed that any effect of the auditory filters themselves on the stimulus phase can be ignored. Indeed, there are only certain aspects of the phase response that have any psychophysically measurable influence. For instance, both the absolute phase and the group delay, proportional to $d\theta/df$, are generally meaningful only in the context of a fixed time reference and cannot be estimated psychophysically (Goldstein 1967; Kohlrausch and Sander 1995). When defining individual auditory filters, the first term that has any perceptual relevance is the phase curvature, $d^2\theta/df^2$, which is the rate of change of group delay as a function of frequency.

Psychoacoustic experiments conducted by Smith et al. (1986) and Kohlrausch and Sander (1995) provided examples of instances where the phase response of the auditory filters cannot be ignored. Both studies used equal-amplitude harmonic-tone complexes with the phases set according to Schroeder (1970), which have come to be known as Schroeder-phase stimuli. Schroeder-phase stimuli have constant phase curvature, implying that the frequency sweep rate is constant (see Kohlrausch and Sander 1995). These stimuli are characterized by very flat temporal envelopes and can be thought of as repeating instantaneously linear rising (Schroeder negative, $m_-$) or falling (Schroeder positive, $m_+$) frequency sweeps. Examples of $m_-$ and $m_+$ stimuli that have been used in Kohlrausch and Sander (1995) and also in the current study are illustrated in Fig. 1. Schroeder-phase complexes have been of interest for psychoacoustic experiments for two reasons. First, $m_+$ and $m_-$ complexes both have an identical power spectrum but produce masked thresholds that can differ by >20 dB, with the $m_+$ stimulus producing the lower threshold. This finding contradicts the power spectrum model of masking (Fletcher 1940; Glasberg and Moore 1990; Patterson 1976), which predicts the same thresholds for $m_+$ and $m_-$ maskers. Second, these complexes provide insight into the phase response of the auditory filters. If frequency-modulated (FM) stimuli, such as an $m_+$ complex and the corresponding (temporally reversed) $m_-$ complex, are passed through a band-pass (auditory) filter, the FM is transferred into an amplitude modulation (AM). For a linear-phase filter, this would result in the same degree of modulation for the two stimuli after filtering. However, it has been proposed that auditory filtering at the level of the basilar membrane (BM) (Rhode and Recio 2000) alters the waveform such that $m_+$ complexes result in a highly modulated or
“peaky” envelope, whereas \( m_- \) complexes produce a less modulated, flatter envelope. This in turn places constraints on the form of the auditory system’s phase response (Kohlrausch and Sander 1995).

Certain semirealistic models of BM filtering (e.g., Giguère and Woodland 1994; Strube 1985) support this view. The output of these models is more highly modulated in response to an \( m_- \) complex than to an \( m_+ \) complex. One reason for the qualitative success of BM models in accounting for the perceptual difference between the \( m_- \) and \( m_+ \) complexes is that the phase response of these models has a negative curvature throughout most of the passband. In the case of the \( m_- \) complex, this negative curvature of the BM filter compensates the positive curvature of the stimulus, leading to a filtered waveform in which the starting phases of all the components come close to coinciding, producing a peaky envelope. Thus when the task is to detect a tone in the presence of the maskers, it appears that the much lower detection thresholds in the case when the task is to detect a tone in the presence of the maskers, whereas

\[
\begin{align*}
\text{amplitude} & \quad 0 \quad 10 \quad 20 \quad 30 \quad 40 \quad 50 \quad 60 \quad 70 \quad 80 \quad 90 \quad 100 \\
\text{time (ms)} & \quad 0 \quad 10 \quad 20 \quad 30 \quad 40 \quad 50 \quad 60 \quad 70 \quad 80 \quad 90 \quad 100
\end{align*}
\]

\[ A: \text{the black lines depict waveforms of harmonic Schroeder-phase complexes (200–2,000 Hz) with 2 choices of starting phase (} m_+, \text{ top;} m_-, \text{ bottom). The gray lines above the maskers indicate the position of the 50-ms signal (1,000 Hz) relative to the 100-ms masker.} \]

\[ B: \text{comparison of the digital waveform of the } m_+ \text{ masker and the corresponding acoustic waveform measured with an artificial ear (2 cc coupler).} \]

In experiment 1, we investigated the AEFs evoked by long-duration tones presented in \( m_+ \) and \( m_- \) Schroeder-phase maskers, with parameters similar to those used in the psychoacoustic study of Kohlrausch and Sander (1995). In experiment 2, the neuromagnetic responses to brief tone pulses presented at different temporal positions within one period of the complex maskers were examined and compared with patterns from loudness experiments, using the same stimuli and derived from the same group of listeners. The results of both experiments were then correlated with simulations of neural activity patterns, using the auditory image model (AIM) of Patterson et al. (1995) that estimates the firing rate in the auditory nerve. These investigations were undertaken in an attempt to better understand 1) how the characteristics of peripheral signal processing are represented at the cortical level of processing in human subjects and 2) to determine the extent to which neuromagnetic responses can be interpreted as an objective correlate of perception.

**Methods**

**Experiment 1:** representation of long-duration tones in complex-tone maskers

**Subjects.** Fourteen subjects for the first experiment (eight female and six male, ages 18 to 43 yr, 13 right-handed) had normal audiometric thresholds and showed no history of peripheral or central hearing disorders. All subjects were familiar with MEG recording sessions and provided informed consent before participating in the experiment, approved by the local ethics committee. Eleven subjects (four female, seven male, ages 24 to 40 yr, ten right-handed) partic-
ipated in the second experiment with nine overlapping in the first experiment.

**STIMULI.** Like Kohlrausch and Sander (1995), we used maskers consisting of equal-amplitude harmonics (range 2 to 20) of a fundamental frequency set at \( f_0 = 100 \) Hz; the starting phase of component \( n \) was assigned according to \( \theta_n = -\pi(n-1)/N \), where \( N \) is the total number of components (in this case 19). Thus the complex is given by: \( \sum_{n=2}^{N} A_n \sin(2\pi f_n t + \theta_n) \). The masker level was 70 dB SPL, corresponding to a level of the individual harmonics of about 57 dB SPL. The complexes derived using the “+” and “−” sign represent, respectively, the Schroeder positive (\( m_+ \)) and the Schroeder negative (\( m_- \)) maskers. The resulting waveforms are shown in Fig. 1. The duration of the maskers was 100 ms and the onsets and offsets were gated using a 10-ms Hanning window.

The signal was a 1,000-Hz tone temporally centered in the masker, as indicated by the gray function in Fig. 1A. Its duration was 50 ms and included a 3-ms Hanning window for the onset and its offset. The signal was added in phase to the spectral component of the respective masker at the signal frequency. Two different signal levels for the masker were chosen in pilot listening experiments such that the signal was clearly detectable at both levels in the case of the masker-plus-signal response and masker-alone response.

Stimuli were generated digitally with a sampling frequency of 44.1 kHz. D/A-conversion was performed using an audio soundcard (RME Audio, Haaimhausen, Germany) connected to a PC. Sounds were presented diotically with a custom-made sound list processor via ER-3 (Etymotic Research, Elk Grove Village, IL) earphones connected to 90-cm plastic tubes and foam earpieces. The interstimulus interval (ISI) was set to 270 ms, measured between the offset and onset of subsequently occurring maskers. The level was adjusted using a Brüel & Kjær sound-level meter type 2203 with an artificial ear (2 cc coupler) 4152. This equipment was attached to a digital oscilloscope (WaveSurfer 424, LeCroy, Chestnut Ridge, NY) and used to investigate the phase and amplitude characteristics of the ER-3 devices. The acoustic waveform of the \( m_- \) masker is illustrated in Fig. 1B (gray) together with the digital waveform (black). The distortion due to the transmission through earphones and tubes is minimal.

**DATA ACQUISITION.** The gradients of the magnetic fields were recorded continuously with a Neuromag-122 whole head system (Elekta Neuromag Oy, Helsinki, Finland) inside a magnetically shielded room (IMEDCO, Hägendorf, Switzerland). Subjects sat in an upright position, viewed a silent movie of their choice, and listened passively to the stimuli. Four coils were attached to the scalp to determine the head position under the dewar during the recordings. Each MEG registration for both the masker-plus-signal condition and hemisphere. Difference waveforms were computed to derive the specific response elicited by the 69- and 57-dB masker signals such that the source waveform of the masker-alone condition without the signal was subtracted from the source waveform of the respective masker-plus-signal condition. To assess the significance of the resulting AEF complex, a permutation test for waveform differences, introduced by Blair and Karniski (1993), was applied. This distribution-free method produces \( P \) values without any assumptions about a particular correlation structure among the waveforms. It can be computed for any number of variables and thus represents a robust alternative to the Hotelling \( T^2 \) test when the number of time points exceeds the number of subjects tested (Picton et al. 2000). The statistic used herein was based on the whole waveform including the P1m maximum of the AEF complex. The output of this procedure is a single multivariate statistic denoted as \( \tilde{t}_{\text{Proc}} \).

Magnetic resonance imaging (MRI) scans, available for 13 of 14 subjects, were obtained using a Siemens Symphony 1.5-T scanner. Three-dimensional reconstructions of the 176 (1-mm voxel) slices were computed using BrainVoyager software (version 4.4, Brain Innovation, Maastricht, The Netherlands). Dipole positions were coregistered onto the individual MRI and then transformed into the standard space of Talairach and Tournoux (1988) to illustrate the location of the generators.

**Experiment 2: period patterns of excitation produced by tone pulses in complex maskers**

**STIMULI.** Kohlrausch and Sander (1995) measured psychoacoustic masking period patterns (the masked threshold of a brief tone pulse, the signal, in the presence of the \( m_+ \) or \( m_- \) masker) as a function of the temporal position of the signal within one period of the masker. In addition to the \( m_+ \) and \( m_- \) complexes, a sine-phase masker with zero starting phase for all masker components (200–2,000 Hz, \( f_0 = 100 \) Hz) was utilized, referred to as the \( m_{00} \) complex. The resulting stimulus resembled a pulse sequence with an interpulse interval of 10 ms, given by the inverse of the fundamental frequency. Portions of the waveforms of the stimuli used in this experiment are illustrated in Fig. 2. The masker level was 70 dB SPL as in experiment 1. The signal was a 5-ms 1,100-Hz tone pulse with 2.5-ms Hanning onset and offset ramps. In contrast to Kohlrausch and Sander, who varied the signal level during their experiment to obtain the corresponding masked threshold, we applied the signal at a fixed level to derive the signal’s specific neuromagnetic response that was affected differently according to the temporal position of the signal in the masker. The level of the signal was fixed at 14 dB relative to the level of the individual masker components (i.e., at 71 dB SPL) such that it was detectable in all signal-masker conditions but close to the masked signal threshold.
for several positions of the signal in the $m_0$ and the $m_-$ maskers (see Fig. 5 in Kohlrausch and Sander 1995). The signal was presented once every 200 ms with a delay of 0, 2, 4, 6, or 8 ms relative to the beginning of the respective masker period. The overall duration of the stimulus was 10 s and the ISI was 1.2 s.

**PSYCHOACOUSTICS.** Psychoacoustic measurements were carried out using the same equipment and stimuli as those used for the MEG registrations. The partially masked loudness produced by the tone embedded in the complex masker was measured. The presence of the masker not only produces a shift in the detection threshold of the signal, from an absolute threshold in quiet to a masked threshold, but also reduces the loudness of the partially masked signal (e.g., Zwicker and Fastl 1999). The partially masked loudness of the signal generally increases as the signal level rises above its masked threshold. Thus we expected that the partial loudness of the signal at the given suprathreshold signal level should be (inversely) related to the masked detection threshold (from Kohlrausch and Sander 1995); i.e., the lower the masked detection threshold for a particular signal-masker condition, the larger should be the corresponding partial loudness sensation. This was investigated for the three masker types ($m_+, m-, m_0$) as a function of the temporal position of the signal in the respective masker.

A two-alternative forced-choice task for paired comparisons of all masker-plus-signal combinations was applied. Subjects were instructed to indicate which signal was perceived as being louder. The Schroeder-phase maskers for the psychoacoustic task were 750 ms in duration and the signal occurred three times: 200, 400, and 600 ms after masker onset. Every combination of paired comparisons, involving three maskers ($m_+, m-, m_0$) and five temporal signal positions (with delays of 0, 2, 4, 6, and 8 ms relative to the onset of the masker period), was presented twice in a different order such that subjects judged 210 pairs overall. A scale for the relative loudness produced by the signal in the masker was derived from the paired comparisons for each subject, using the Bradley–Terry–Luce (BTL) method (David 1988), which assumes that perceived loudness can be ordered on a single scale.

**MEG RECORDINGS.** The effects of phase-sensitive cochlear processing in neural representations at the cortical level were studied using auditory evoked fields. These were recorded in response to the same stimuli as those in the psychoacoustic experiments. The patterns of excitation, produced by the signal added to the different maskers ($m_+, m-, m_0$) at different positions within a masker period, are referred to as neuromagnetic period patterns. Recording settings similar to those reported in experiment 1 were applied with the one exception that about 1,100 single sweeps, from $-50$ to $+250$ ms relative to the stimulus onset, were averaged for each stimulus condition. MEG registrations, obtained separately for each masker condition, were taken for about 30 min in each session.

**DATA ANALYSIS.** The fitting procedure for the source analysis was nearly the same as that in the first experiment, with the exception that the specific response evoked by a signal was fitted with a slightly different procedure. A BESA model with one equivalent dipole in each hemisphere was computed for each masker condition and fitted to the experimental condition that exhibited the largest response evoked by a signal (i.e., the 0-ms-delay condition for the $m_+$ and $m_-$ maskers, and the 4-ms-delay condition for the $m_0$ masker). This procedure resulted in three slightly different fits for each subject and masker condition, as shown in Fig. 3. A common reliable source model for each subject was derived by averaging the coordinates and orientations of the three independent source models. These mean parameters were used in the final model for all 15 experimental conditions according to masker type ($m_+, m-, m_0$) and signal position (0, 2, 4, 6, and 8 ms). The above-cited procedure, applied for each subject, assumed that the P1m response is evoked by the same generators within Heschl’s gyrus; i.e., the location and orientation of the equivalent dipole remained constant for the signal presented at the different temporal positions.

An ANOVA (general linear model, SAS software, version 9) with repeated measurements including the factors masker and signal position.
tion was used to investigate the main and interaction effects of the P1m maxima elicited by the different masking types and relative delays of the signal.

**Simulation of auditory-nerve activity patterns**

A nonlinear model of cochlear processing was used to consider the effects of BM filtering and transformation into auditory-nerve (AN) activity for the stimuli of the present study. The auditory image model (AIM) of Patterson et al. (1995) was used to simulate the fine-grain spectrotemporal information in the auditory nerve. The first stage of this model consists of a one-dimensional transmission-line filterbank that accounts for cochlear hydrodynamics (Giguère and Woodland 1994). The output of this stage simulates basilar-membrane motion. The simulations in the present study were computed with a transmission line consisting of 500 sections covering a frequency range from 100 to 6,000 Hz. The tuning of the local segments of the basilar membrane was set to $Q_n = 8$, as suggested by Carlyon and Datta (1997). They used the same modeling framework to account for the effects of the signal level, the number of components, and the phase of flanking components in psychoacoustic masking period patterns using the same type of maskers. The output of the transmission line was then converted into the neural activity pattern (NAP) using a hair-cell simulator with one hair cell per filterbank channel (Meddis 1988). A medium and a high spontaneous-rate fiber type were applied for the hair cell of each simulated channel. As in Berlin (1984), the excitation patterns of the medium and high spontaneous-rate fibers were weighted respectively by 0.35 for the medium spontaneous rate and 0.65 for the high spontaneous rate before adding the data for both fiber types.

**RESULTS**

**Experiment 1: representation of long-duration tones in complex-tone maskers**

**MEG DATA.** It was possible to consistently fit a stable spatio-temporal model with one equivalent dipole in each auditory cortex for all 14 subjects. The projection of the equivalent dipoles onto an averaged map of the sulcal borders of Heschl’s gyrus (Schneider et al. 2005) showed dipole sources located bilaterally close to the border of the medial and lateral Heschl’s gyrus (Fig. 3). Figure 4 displays individually the grand average source waveforms obtained with the $m_+$ condition (Fig. 4A) and the $m_-$ conditions (Fig. 4B).

The results for the $m_+$ condition (Fig. 4A) are as follows: The masker-alone condition (black curve, left) evoked a Pam followed by a prominent P1m and then a small N1m at about 130 ms. The addition of the more intense signal (69 dB SPL), indicated by the light gray curve in the same panel, generated a clear additional response with a peak at about 60 ms relative to the signal onset. The signal-related response, derived by subtracting the waveform generated by the $m_+$-masker alone from that produced by the $m_+$-masker plus signal, is shown in the corresponding right panel. The waveform exhibited a significant positive plateau starting at about 60 ms after signal onset and lasting until about 150 ms ($t_{\text{sum}} = 227.2$, $P < 0.01$). For the weaker signal (57 dB; indicated by the light gray curve), a smaller response was observed with a delayed peak where the largest positive response occurred near 100 ms. Permutation tests showed that this slow positive AEF represented a significant deflection ($t_{\text{sum}} = 555.2$, $P < 0.001$).

For the $m_-$ condition (Fig. 4B), the source waveforms showed a steeper slope of the initial Pam activation than that for the $m_+$ condition in response to the masker alone. The whole waveform exhibited a strong ripple with a periodicity of 10 ms corresponding to that of the masker waveform. The corresponding signal-related responses, i.e., the differences between the AEFs obtained with the $m_-$ masker plus signal and those with the masker alone, are shown in Fig. 4B (right).

The more intense signal (69 dB) evoked a significant positivity ($t_{\text{sum}} = 227.9$, $P < 0.01$). For the weaker signal, the specific response resulted in a difference source waveform with a much smaller magnitude than that obtained in the $m_+$ condition, albeit with a significant deflection ($t_{\text{sum}} = 310.1$, $P < 0.01$).

The direct comparison of the signal-related AEF (Fig. 4A, right vs. Fig. 4B, right) between the masker conditions did not reveal a significant difference for the more intense signal ($t_{\text{sum}} = 85.8$, ns). However, a significant difference was found for the weaker signal such that the signal-related AEF was larger in the $m_+$ condition than that in the $m_-$ condition ($t_{\text{sum}} = 241.4$, $P < 0.01$).

Finally, we compared the responses obtained with the two maskers alone (Fig. 4C). The difference of the response waveforms is shown in the corresponding right panel. The permutation test showed a highly significantly larger response for the $m_+$ masker than that for the $m_-$ masker in the interval from 30 to 150 ms that covers the whole positive deflection ($t_{\text{sum}} = 308.0$, $P < 0.0001$).

**SIMULATED AN ACTIVITY PATTERNS.** Figure 5A shows the internal representation of the four masker-plus-signal combinations after nonlinear basilar-membrane processing (Giguère and Woodland 1994) through a filter tuned to a signal frequency of 1,000 Hz. The gray functions show the filtered masker-alone responses and the black functions indicate the filtered representations of the sum of masker and signal. The filtered $m_+$ and $m_-$ maskers have very different shapes, especially in terms of their “peakiness,” as a consequence of the dispersive characteristics of the BM transformation. This can be seen most clearly in the first periods of the filtered maskers (in the absence of the signal). The $m_+$ masker is modulated more strongly in amplitude than the $m_-$ masker. This is also reflected in the neural activity patterns (NAPs) shown in Fig. 5B. In the framework of the cochlea model, the addition of the signal produces a stronger change of neural activity in the case of the $m_+$ masker (where the signal effectively “fills” the valleys of the filtered masker representation) than in the case of the $m_-$ masker.

Figure 5C shows the neural summary pattern at the AN level, integrated across the nominal frequency range between 200 and 2,000 Hz. The AN summary activity evoked by the $m_-$ masker alone (i.e., for the first and last 20 ms) shows a rippled structure. This result can be explained by the superposition of the neural activity across frequencies. The upward sweep ($m_-$ masker) partially compensates the traveling-wave delay across frequency, observed when the excitation by a (transient) stimulus progresses apically along the BM. This results in a relatively strong response amplitude for each period of the stimulus. In contrast, the summary NAP obtained with the $m_+$ masker shows a less pronounced 10-ms periodicity. Instead, the $m_+$ pattern exhibits much stronger fluctuations within a stimulation period due to the more peaky responses of the individual channels, as exemplified by the channel at the resonance frequency of 1,000 Hz in Fig. 5A. Since the peaks
stemming from different peripheral channels occur at different times, their superposition results in a rippled fine structure of the summed waveform. At the high signal level, the signal-related change of excitation in the temporal portion where the signal is added to the masker is substantial in both masker conditions (\(m_+\) and \(m_-\)). At the lower signal level, the responses are correspondingly weaker. However, the signal-related response appears to be stronger for the \(m_+\) masker than that for the \(m_-\) masker. In the latter case, the signal contributes only within very short portions of activity at the beginning of each masker period.

Some of the characteristics in the summed NAPs after cochlear processing (Fig. 5C) are consistent with our measured neuromagnetic responses (from Fig. 4). The simulated periodic 10-ms modulation of the activity produced by the \(m_-\) masker alone corresponds to the periodicity in the neuromagnetic response from Fig. 4B. Importantly, the signal-related responses in the NAP simulations are consistent with the corresponding signal-related neuromagnetic source waveforms, where the excitation due to the (lower-level) signal was larger in the presence of the \(m_+\) masker than that in the presence of the \(m_-\) masker (Fig. 4, A and B, right panels).

**Experiment 2: correlation of neuromagnetic period patterns, partial loudness patterns, and simulated auditory-nerve activity**

**MEG DATA.** The grand average AEF source waveforms are shown in Fig. 6. The responses evoked by the maskers alone (thin black curves) indicate the baseline conditions for the \(m_+\) (Fig. 6, top), \(m_-\) (Fig. 6, middle), and \(m_0\) masker (Fig. 6, bottom). As discussed in the first experiment, the responses to the \(m_-\) masker show a pronounced ripple structure. The remaining curves in each panel display the responses to the different masker-plus-signal combinations. The addition of the transient 5-ms signal to the masker produced a large
positive response that includes a Pam at about 30 to 40 ms in almost all conditions. This is followed by a large P1m peaking at about 80 ms, found for all three maskers and all temporal positions of the signal in the masker. The P1m peak value of the response waveform varies as a function of the temporal position of the signal in the masker.

The mean peak amplitudes, averaged across subjects, are shown in Fig. 7A for the different masker-signal configurations. For the \( m_+ \) and \( m_- \) maskers, the resulting neuromagnetic period patterns show similar shapes but are shifted in magnitude from one another. There is a significant main effect of masker type (see Table 1) in the \( m_+ \) and \( m_- \) data and a nonsignificant interaction between masker type and signal delay. The largest response amplitudes for these two masker conditions were evoked when the signal was presented at delays of 0 and 8 ms, whereas the smallest response was found at a delay of 4 ms. In contrast, the results for the \( m_0 \) masker condition show a different pattern, with the minimum at a signal delay of 0 ms and a broad plateau for signal positions of \( \geq 4 \) ms. Here, the interaction masker \( \times \) phase was significant for the \( m_+ \) and \( m_0 \) as well as the \( m_- \) and \( m_0 \) conditions.

**Psychophysical Data.** The perceptual data were obtained by a pairwise loudness comparison with the same stimuli as those used in the MEG recordings. The resulting pattern of partially masked loudness values, based on the BTL procedure (David 1988), is shown in Fig. 7B. The general shape is similar to the neuromagnetic period pattern (Fig. 7A). The loudness produced by the signal presented in the \( m_0 \) masker showed an almost inverted pattern relative to that obtained with the Schroeder-phase maskers. The highest loudness values for the \( m_0 \) masker were found for signal delays at 4 to 8 ms, whereas a minimum was observed at 0 ms.

**Simulated Activity Patterns.** The specific excitation produced by the signal was calculated by subtracting the spike rate evoked by the masker alone from that of the masker plus signal. This signal-related excitation is depicted in Fig. 8. For each masker condition, the additional activation was determined for the five different positions of the signal in the masker. The thin lines represent the activation produced by the maskers alone, summed across channels from 550 to 2,200 Hz. The shaded areas above the lines indicate additional activity from the increase in spike rate evoked by the signal under the various conditions. Integration of the shaded areas (i.e., the

![Figure 5](http://jn.physiology.org/)

**Fig. 5.** Simulated basilar-membrane motion and neural activity patterns (NAPs) using a one-dimensional transmission-line model. A: basilar-membrane response at 1,000 Hz to the masker alone (gray) and the masker plus signal (black). B: NAP at 1,000 Hz for the same stimulus conditions as in A. C: summary NAP in the range from 200 to 2,000 Hz.

The mean peak amplitudes, averaged across subjects, are shown in Fig. 7A for the different masker-signal configurations. For the \( m_+ \) and \( m_- \) maskers, the resulting neuromagnetic period patterns show similar shapes but are shifted in magnitude from one another. There is a significant main effect of masker type (see Table 1) in the \( m_+ \) and \( m_- \) data and a nonsignificant interaction between masker type and signal delay. The largest response amplitudes for these two masker conditions were evoked when the signal was presented at delays of 0 and 8 ms, whereas the smallest response was found at a delay of 4 ms. In contrast, the results for the \( m_0 \) masker condition show a different pattern, with the minimum at a signal delay of 0 ms and a broad plateau for signal positions of \( \geq 4 \) ms. Here, the interaction masker \( \times \) phase was significant for the \( m_+ \) and \( m_0 \) as well as the \( m_- \) and \( m_0 \) conditions.

**Psychophysical Data.** The perceptual data were obtained by a pairwise loudness comparison with the same stimuli as those used in the MEG recordings. The resulting pattern of partially masked loudness values, based on the BTL procedure (David 1988), is shown in Fig. 7B. The general shape is similar to the neuromagnetic period pattern (Fig. 7A). The loudness produced by the signal presented in the \( m_0 \) masker showed an almost inverted pattern relative to that obtained with the Schroeder-phase maskers. The highest loudness values for the \( m_0 \) masker were found for signal delays at 4 to 8 ms, whereas a minimum was observed at 0 ms.

**Simulated Activity Patterns.** The specific excitation produced by the signal was calculated by subtracting the spike rate evoked by the masker alone from that of the masker plus signal. This signal-related excitation is depicted in Fig. 8. For each masker condition, the additional activation was determined for the five different positions of the signal in the masker. The thin lines represent the activation produced by the maskers alone, summed across channels from 550 to 2,200 Hz. The shaded areas above the lines indicate additional activity from the increase in spike rate evoked by the signal under the various conditions. Integration of the shaded areas (i.e., the

![Figure 6](http://jn.physiology.org/)

**Fig. 6.** Grand average source waveforms of a 2-dipole model based on the fit of the middle latency-onset response. The different curves represent the responses to the masks alone (thin black lines) and to the masks plus the signal, with signal delay as the parameter. The results for the \( m_+ \), \( m_- \), and \( m_0 \) maskers are shown in the top, middle, and bottom panels, respectively.
TABLE 1. Results of repeated-measures ANOVA to assess the significance of neuromagnetic masking effects due to different masker types and temporal positions of the signal in the masker in experiment 2

<table>
<thead>
<tr>
<th>Masker Condition</th>
<th>Source of Variation</th>
<th>F-Statistic</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m_+ - m_-$</td>
<td>masker</td>
<td>$F_{(1,10)} = 40.37$</td>
<td>$P &lt; 0.001$</td>
</tr>
<tr>
<td></td>
<td>phase</td>
<td>$F_{(4,40)} = 7.44$</td>
<td>$P &lt; 0.001$</td>
</tr>
<tr>
<td></td>
<td>masker $\times$ phase</td>
<td>$F_{(4,40)} = 1.16$</td>
<td>ns</td>
</tr>
<tr>
<td>$m_+ - m_0$</td>
<td>masker</td>
<td>$F_{(1,10)} = 6.03$</td>
<td>$P &lt; 0.05$</td>
</tr>
<tr>
<td></td>
<td>phase</td>
<td>$F_{(4,40)} = 2.31$</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>masker $\times$ phase</td>
<td>$F_{(4,40)} = 14.03$</td>
<td>$P &lt; 0.001$</td>
</tr>
<tr>
<td>$m_- - m_0$</td>
<td>masker</td>
<td>$F_{(1,10)} = 4.89$</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>phase</td>
<td>$F_{(4,40)} = 5.17$</td>
<td>$P &lt; 0.01$</td>
</tr>
<tr>
<td></td>
<td>masker $\times$ phase</td>
<td>$F_{(4,40)} = 12.42$</td>
<td>$P &lt; 0.001$</td>
</tr>
<tr>
<td>$m_+ - m_- - m_0$</td>
<td>masker</td>
<td>$F_{(2,20)} = 12.93$</td>
<td>$P &lt; 0.001$</td>
</tr>
<tr>
<td></td>
<td>phase</td>
<td>$F_{(4,80)} = 3.30$</td>
<td>$P &lt; 0.05$</td>
</tr>
<tr>
<td></td>
<td>masker $\times$ phase</td>
<td>$F_{(8,80)} = 10.03$</td>
<td>$P &lt; 0.001$</td>
</tr>
</tbody>
</table>

ns, nonsignificant.

overall change in neural activity due to the addition of the signal) leads to the corresponding values shown in Fig. 7C. The resulting patterns for the different maskers and signal positions are similar to those obtained from the neuromagnetic recordings (Fig. 7A) and the psychoacoustic partial loudness data (Fig. 7B). The locations of the maxima and the minima of the patterns as well as the shift between them for the $m_+$ and $m_-$ maskers are also similar.

FIG. 7. A: neuromagnetic period pattern as given by the maximum values of the source waveforms from Fig. 6. The bars indicate the SE. B: psychoacoustic loudness patterns derived from a 2-alternative forced-choice task under all stimulus conditions (Bradley–Terry–Luce scale). C: simulated auditory nerve (AN) period patterns based on the summary activity patterns shown in Fig. 8.

FIG. 8. Simulation of the summary AN activity patterns (NAPs) for the stimuli used in experiment 2. The gray areas show the increase of excitation evoked by the 1,100-Hz signal (summed across the frequency channels from 550 to 2,200 Hz).

COMPARISON OF SIMULATED AN ACTIVITY, AEF, AND PERCEPTUAL LOUDNESS DATA. A different view on the relationship between the three measures is provided in Fig. 9, which shows the scatterplots of the pairwise relationships among the neuromagnetic P1m, the psychoacoustic loudness judgments, and the simulated AN activity patterns. Each panel includes 15 data points reflecting the mean values for the subjects obtained for the five different temporal positions of the signal presented in the three maskers. The loudness results correlate significantly with the P1m magnitude of the neuromagnetic cortical response (Fig. 9A, $r = 0.91, P < 0.05$) and with the simulated AN spike activity, summed across frequencies (Fig. 9B, $r = 0.84, P < 0.05$). The measured neuromagnetic response was also significantly correlated with the simulated AN activity (Fig. 9C, $r = 0.87, P < 0.05$).

DISCUSSION

In this study, we have compared neuromagnetic responses with perceptual data and simulated cochlear activity patterns to
signal onset in the P1m from auditory cortex that started about 70 ms after Heschl’s gyrus. Previous studies have shown that the N1m is strongly reduced for ISIs <0.5 s (Hari et al. 1982; Onitsuka et al. 2000) and not consistently observed for ISIs <300 ms (Carver et al. 2002), whereas the P1m is only slightly attenuated at an ISI of 200 ms and can still be observed for ISIs of ≤100 ms (Gutschalk et al. 2004).

In both experiments of this study, the signal-related neuro-magnetic responses were consistent with the perceptual data and with changes of excitation in the neural activity pattern at the output of cochlear processing. The difference waveforms of the responses in experiment 1 (Fig. 4, A and B) showed a plateau-like positive deflection for the more intense signal (69 dB SPL) for both masker types. At this level, the two signal-related responses did not differ in magnitude from each other. For the lower signal level (57 dB SPL), the responses were weaker and increased more gradually with time (and also decreased more gradually after signal offset). Here, the signal-related response was larger for the m+ than that for the m− condition. This observation was consistent with the simulations of the AN activity patterns (Fig. 5). These results are also consistent, at least qualitatively, with the perceptual masking results from Kohlrausch and Sander (1995) that demonstrated that the signal is masked more effectively by the m+ complex than by the m− masker.

A more detailed and direct investigation of the relation between AEF and perception was considered in our experiment 2. Regarding the correlation with the perceptual data, it is interesting to relate the loudness results from this experiment to the masked threshold data provided in Kohlrausch and Sander (1995). Figure 10 shows a replot of their masked detection data (their Fig. 5), using the same axes as in Fig. 7 of the present study. The masked detection threshold in Fig. 10 is shown as a function of the temporal position of the signal within (one period of) the masker. For the m0 masker, the peak in the masked period pattern occurs at the signal position close to 0 ms. The reason for this is that the envelope of this masker has a peak close to 0 ms (see Fig. 2) and thus makes it difficult to detect the signal. For the m+ and m− maskers, the maxima of the period patterns are at a signal position at which the instantaneous frequency of the masker coincides with the signal frequency. Since the signal frequency is in the spectral center of the complex, this time point occurs approximately in the temporal center of the period. Thus compared with the maxima in the period pattern for the m0 complex, the maxima for the Schroeder-phase complexes are temporally shifted by

study the representation of cochlear filtering characteristics at a cortical level of processing in human subjects. We used specific broadband maskers that differ largely in their effectiveness to mask a tonal signal. These masking differences have in previous studies been associated with properties of the complex transfer function of the auditory filters tuned at the signal frequency. Specifically, we examined the changes of activity in the neuromagnetic responses due to the addition of high-frequency maskers, and with changes of excitation in the neural activity pattern at the output of cochlear processing. The difference waveforms of the responses in experiment 1 (Fig. 4, A and B) showed a plateau-like positive deflection for the more intense signal (69 dB SPL) for both masker types. At this level, the two signal-related responses did not differ in magnitude from each other. For the lower signal level (57 dB SPL), the responses were weaker and increased more gradually with time (and also decreased more gradually after signal offset). Here, the signal-related response was larger for the m+ than that for the m− condition. This observation was consistent with the simulations of the AN activity patterns (Fig. 5). These results are also consistent, at least qualitatively, with the perceptual masking results from Kohlrausch and Sander (1995) that demonstrated that the signal is masked more effectively by the m+ complex than by the m− masker.

A more detailed and direct investigation of the relation between AEF and perception was considered in our experiment 2. Regarding the correlation with the perceptual data, it is interesting to relate the loudness results from this experiment to the masked threshold data provided in Kohlrausch and Sander (1995). Figure 10 shows a replot of their masked detection data (their Fig. 5), using the same axes as in Fig. 7 of the present study. The masked detection threshold in Fig. 10 is shown as a function of the temporal position of the signal within (one period of) the masker. For the m0 masker, the peak in the masked period pattern occurs at the signal position close to 0 ms. The reason for this is that the envelope of this masker has a peak close to 0 ms (see Fig. 2) and thus makes it difficult to detect the signal. For the m+ and m− maskers, the maxima of the period patterns are at a signal position at which the instantaneous frequency of the masker coincides with the signal frequency. Since the signal frequency is in the spectral center of the complex, this time point occurs approximately in the temporal center of the period. Thus compared with the maxima in the period pattern for the m0 complex, the maxima for the Schroeder-phase complexes are temporally shifted by

FIG. 9. Scatterplots including linear regressions to illustrate the close correspondence among AN excitation, neuromagnetic responses (AEFs), and perceived loudness (BTL) under all stimulus conditions. All correlations were significant (P < 0.05).

FIG. 10. Masked thresholds of a 5-ms signal as a function of the onset time difference between masker and signal. Adapted from Kohlrausch and Sander (1995; Fig. 5).
half a period (i.e., 5 ms), as can be seen in Fig. 10. The function for the $m_-$ masker, however, is shifted toward higher threshold values and is somewhat flatter than the pattern for the $m_+$ masker. The differences between these two patterns have been attributed to the dispersive properties of the peripheral filter tuned to the signal frequency (Kohlrausch and Sander 1995; Smith et al. 1986).

With respect to the loudness experiments of this study, we expected that those signal-masker combinations that produced the lowest masked signal thresholds (in Fig. 10) should evoke the largest (partial) loudness of the signal. Indeed, the loudness values (Fig. 7B) are approximately inversely related to the masked thresholds in Fig. 10. For example, for a given temporal position of the signal in the masker, such as 4 ms, the $m_0$ masker produced the lowest signal detection threshold, whereas the $m_-$ masker produced the largest value. In the corresponding loudness data from Fig. 7B, the $m_0$ masker led to the largest loudness of the signal, whereas the $m_-$ masker produced the lowest loudness. Accordingly, the loudness patterns for the individual masker types have shapes that are approximately inverted versions of the masked threshold patterns from Fig. 10, suggesting that the two patterns may underlie similar processes. However, the agreement is only qualitative and it can be expected that the shape of masked period patterns and (partial) loudness patterns depend on the stimulation level since cochlear filtering is known to nonlinearly depend on level (e.g., Ruggiero et al. 1997).

A relationship between the P1m and perception has earlier been observed in a study of auditory stream segregation (Gutschalk et al. 2005), where the P1m was related to the strength of perceived segregation of one stream of tones from another one. Comparing stimuli of very different complexity, Lütkenhöner et al. (2006) provided evidence that activation strength in the auditory periphery correlates with the P1m magnitude. Moreover, Chait et al. (2004) demonstrated that the P1m was relatively independent of task demands and perturbation. Thus the P1m appeared to be well suited to study the relationship between peripheral filtering, auditory cortex function, and perception, as confirmed here by the high correlation found between summed AN activity and the P1m magnitude (Fig. 9C).

Another interesting observation was made in the present study with respect to the masker-alone responses. The masker-related response amplitude was larger for the $m_-$ masker than for the $m_+$ masker (Fig. 4C). This is consistent with the findings of Recio (2001) who recorded larger firing rates in AN fibers and neurons of the ventral cochlear nucleus (VCN) of chinchillas for their $m_-$ masker than for their $m_+$ masker. Our results are also consistent with those of Rupp et al. (2002) who showed that middle-latency responses reflect the amount of synchronization along the basilar membrane. Their responses, obtained with upward sweeping chirps of a sweeping rate (Dau et al. 2000) different from that used here, also showed larger amplitudes than those obtained with a temporally reversed (downward sweeping) chirp. In Dau et al. (2000), Rupp et al. (2002), and related studies (e.g., Fobel and Dau 2004; Junius and Dau 2005; Stürzebecher et al. 2006), the approach was different from that in the present study. In their studies, the goal was to maximize neural synchronization across frequency that was achieved using a broadband stimulus that attempts to compensate for cochlear travel-time differences. Such a stimulus must be rising in instantaneous frequency with time since low-frequency components require more time to reach their place of maximum displacement (near the apex of the cochlea) than high-frequency components (close to the base). The upward-sweeping $m_+$ masker of the present study, even though much narrower in bandwidth and reflecting a linear sweeping rate [in contrast to the nonlinear sweeping rate in Dau et al. (2000)], produces a larger overall response amplitude than the $m_-$ masker since it generates a larger amount of synchronization in the excited frequency region. This, in turn, is consistent with recent results by Mauermann and Hohmann (2007) on loudness perception where the listeners required a 6-dB higher stimulation level for $m_+$ complexes to match the loudness of the $m_-$ complexes, a result that shows that phase-sensitive cochlear processing can affect loudness perception and should be incorporated in loudness models.

Physically, phase dispersion in the transfer function of individual cochlear filters in the form of an upward frequency “glide” (e.g., de Boer and Nuttall 1997) in the impulse responses is equivalent to (or a consequence of) the occurrence of the traveling wave along the basilar membrane. It is not possible to create a stimulus that compensates both for travel-time differences across frequency and for the phase curvature in the individual auditory filters. Thus although the $m_-$ masker led to larger signal-related responses, which were in the focus of the present study, the $m_+$ stimulus produced the larger masker-alone response.

The results presented in this study suggest that neuromagnetic responses, and in particular the middle-latency AEFs, might be beneficially used as an objective correlate of the behavioral performance of human subjects in auditory masking conditions. They might also be helpful for elucidating models of cochlear processing in humans through comparison of neuromagnetic recordings with model predictions. The neuromagnetic responses considered here might also be interesting as an objective tool in hearing research involving patients that cannot participate actively in listening experiments such as newborns and small children. Recent results in psychoacoustic masking studies with hearing-impaired listeners, using Schroeder-type tone-complex masksers, showed much flatter masked period patterns compared with those of normal-hearing listeners (Oxenham and Dau 2004; Summers 2000; Summers and Leek 1998). The variation in these patterns has been suggested to reflect the state of hearing in human listeners.

ACKNOWLEDGMENTS

We thank I. Nellen for helpful discussions and K. Krumbholz and one anonymous reviewer for very helpful criticism and suggestions.

GRANTS

This work was supported by Deutsche Forschungsgemeinschaft Grant Ru-652/1-3 and Danish Research Foundation Grant 26-04-0265.

REFERENCES

Carver FW, Fuchs A, Jantzen KJ, Kelso JAS. Spatiotemporal analysis of the neuromagnetic response to rhythmic auditory stimulation: rate depen-