Research paper

Evidence of pitch processing in the N100m component of the auditory evoked field

Annemarie Seither-Preisler [a,*], Roy Patterson [b], Katrin Krumbholz [c], Stefan Seither [a], Bernd Lütkenhöner [a]

[a] *Department of Experimental Audiology, ENT Clinic, Münster University Hospital, Kardinal von Galen-Ring 10, D-48129 Münster, Germany*
[b] *Centre for the Neural Basis of Hearing, Department of Physiology, University of Cambridge, Downing Street, CB2 3EG Cambridge, UK*
[c] *MRC Institute of Hearing Research, University Park, Nottingham NG7 2RD, UK*

Abstract

The latency of the N100m component of the auditory evoked field (AEF) is sensitive to the period and spectrum of a sound. However, little attention was paid so far to the wave shape at stimulus onset, which might have biased previous results. This problem was fixed in the present study by aligning the first major peaks in the acoustic waveforms. The stimuli were harmonic tones (spectral range: 800–5000 Hz) with periods corresponding to 100, 200, 400, and 800 Hz. The frequency components were in sine, alternating or random phase. Simulations with a computational model suggest that the auditory-nerve activity is strongly affected by both the period and the relative phase of the stimulus, whereas the output of the more central pitch processor only depends on the period. Our AEF data, recorded from the right hemisphere of seven subjects, are consistent with the latter prediction: The latency of the N100m depends on the period, but not on the relative phase of the stimulus components. This suggests that the N100m reflects temporal pitch extraction, not necessarily implying that the underlying generators are directly involved in this analysis.
© 2006 Elsevier B.V. All rights reserved.

Keywords: N100m latency; AEF; Pitch; Relative phase; AIM

1. Introduction

The auditory evoked field (AEF) measured by magneto-encephalography (MEG) basically represents cortical responses to afferent inputs from earlier stages of the auditory pathway. A stimulus-dependence of a given AEF component may originate from any stage in the processing hierarchy, from the auditory periphery up to the site where the AEF is generated. It is therefore desirable to identify the origin of systematic AEF variations. Experiments combined with auditory model simulations may provide a promising approach, as they allow establishing co-variations of stimulus representations at different functional levels with cortical AEF-patterns.

One of the best-investigated AEF components is the N100m deflection, peaking about 100 ms after the onset of a sound. The response arises from multiple sources in auditory cortex (Näätänen and Picton, 1987; Sams et al., 1993; Loveless et al., 1996), with a center of activity in planum temporale (Lütkenhöner and Steinsträter, 1998). The structures associated with the N100m receive inputs from cortical and subcortical areas (Eggermont

---

Abbreviations: AEF, auditory evoked field; AI, auditory image; AIM, auditory image model; ALT, complex tone with harmonics in alternating phase; ANOVA, analysis of variance; BMM, basilar membrane motion; EEG, electroencephalography; ERB, equivalent rectangular bandwidth; F0, fundamental frequency; MEG, magnetoencephalography; NAP, neural activity pattern; POR, pitch onset response; RAND, complex tone with harmonics in random phase; RMS, root-mean-square (effective value); SIN, complex tone with harmonics in sine phase; SOA, stimulus onsets asynchrony (time interval between stimulus onset)

* Corresponding author. Tel.: +49 251 83 56817; fax: +49 251 56882.
E-mail addresses: preisler@uni-muenster.de (A. Seither-Preisler), roy.patterson@mrc-cbu.cam.ac.uk (R. Patterson), katrin@ihr.mrc.ac.uk (K. Krumbholz).

and Ponton, 2002). As the N100m and its electroencephalographic counterpart, the N100, are sensitive to the intensity (Rapin et al., 1966; Beagley and Knight, 1967) and the rise-time of the sound (Biermann and Heil, 2000), it is often regarded as an energy-onset response. However, there is accumulating evidence that the deflection is also sensitive to other stimulus features, such as spectral composition (Roberts et al., 2000; Lütkenhöner et al., 2001; Seither-Preisler et al., 2003), pure tone frequency (Jacobson et al., 1992; Roberts and Poeppel, 1996; Stufflebeam et al., 1998; Lütkenhöner et al., 2001), and the fundamental frequency of harmonic tones (Ragot and Lepaul-Ercole, 1996; Ragot and Crottaz, 1998; Crottaz-Herbette and Ragot, 2000). It is not clear yet to what extent the N100m is a function of stimulus representations in the auditory periphery or of a more central processing. To get more insight, we compared cortical AEFs with simulated representations of complex tones at the level of the basilar membrane, the auditory nerve, and a central pitch extraction stage. While our stimuli had different periods and phase spectra, the frequency band was fixed, to avoid differences in cochlear travelling wave delay.[1] In addition, we aligned the first major peaks in the stimulus waveform, to avoid confounds related to variations in the effective onset time.

2. Materials and methods

2.1. Stimulus design

2.1.1. Problems with conventional stimuli

Most auditory stimuli differ in the cochlear travelling wave delays they produce before optimally stimulating the auditory nerve. According to Patterson (1994), the latencies of the travelling wave maxima for a 100 Hz and a 5 kHz pure tone differ by about 12 ms. Even if broadband stimuli are filtered into the same frequency region, to match them for travelling wave delay, differences in phase and period may cause latency variations of the first major peak in the signal. This will, in turn, affect the latency of subsequent neural responses. The problem is illustrated in Fig. 1, which shows the first one and a half cycles of a complex tone consisting of the first six harmonics of a fundamental frequency (F0) of 800 Hz. The peaks in the composite waveform (panel (b)) result from the superposition of the ascending slopes of the component sinusoids (panel (a)). The very first peak (at 0.09 ms in the present example) rises immediately after stimulus onset, regardless of F0. The latency of the second peak is a function of F0, as illustrated in Fig. 2. The two panels in Fig. 2 show the waveforms of harmonic tones with frequency components between 800 to 5000 Hz and periods corresponding to 100 Hz (a) and 800 Hz (b). The second major peak has a latency of about 10 ms when the period corresponds to 100 Hz, and of only 1.34 ms (0.09 + 1.25 ms; cf., Fig. 1b) when the period corresponds to 800 Hz. To avoid spectral splatter associated with rapid energy transitions, a windowing function is commonly applied to stimulus onset and offset. In Fig. 2, we assume an onset ramp of 5 ms (shaded areas). For stimulus (a), the onset-ramp would eliminate the very first peak of the signal, leaving the next peak
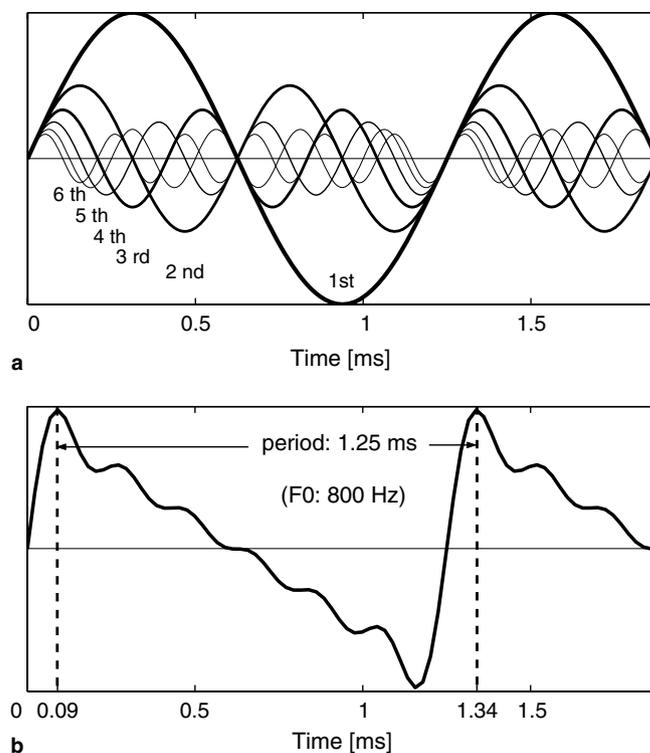
Fig. 1. First one and a half cycles of a complex tone with a fundamental frequency (F0) of 800 Hz. (a) The first 6 harmonics. (b) Resulting composite waveform.
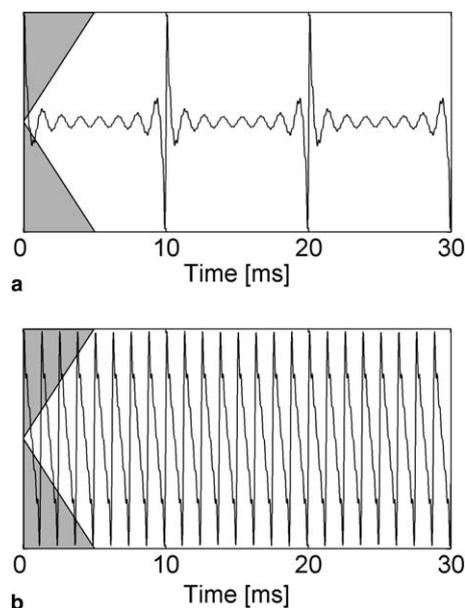

Fig. 2. Waveforms of harmonic tones with F0s of 100 Hz (a) and 800 Hz (b). The spectral region is 800–5000 Hz in both cases. For stimulus (a) the first peak occurs around 10 ms, whereas for stimulus (b) small peaks arise already around 2 ms and a maximal amplitude is reached after 5 ms. The shaded areas indicate signal portions that would be removed by a typical onset ramp of 5 ms.

at about 10 ms. For stimulus (b), some small peaks would arise after <2 ms, reaching their maximal amplitude at the end of the onset ramp at 5 ms. In view of the fact that the typical latency variations observed

in AEF-studies are of the order of milliseconds, such stimulus-induced effects represent a problematic source of variability that should be controlled.

2.1.2. Set of stimuli used in this study

The stimuli were complex tones with harmonics in the frequency range between 800 and 5000 Hz, and periods corresponding to 100, 200, 400, and 800 Hz. The frequency spectra of the stimuli are displayed in the upper row of Fig. 3. The F0 itself was missing in the 100, 200 and 400-Hz stimuli, but present in the 800-Hz stimulus. In order to achieve a smooth and natural timbre, the amplitudes of the harmonics decreased by 6 dB per octave relative to F0. To account for the transfer characteristic of the transmission system (variation of ±14 dB in the range between 100 and 4000 Hz), the amplitude of each spectral component was multiplied by a factor compensating for the system's frequency-specific loss in sound pressure level. By this means, the original spectral profile of the stimuli was preserved. The compensation is not shown in Fig. 3.

For each of the four F0s, the tones were generated either in sine (SIN), alternating (ALT), or random (RAND) phase, yielding 12 stimuli. In SIN tones, all harmonics undergo their first zero-crossing at time 0. In ALT tones, the zero-crossings of the even harmonics are shifted by +90° relative to those of the odd harmonics. In RAND tones, the phase of each component is randomly drawn from a uniform distribution between 0 and 2π. All tones were presented with the same root-mean-square (RMS) level. To eliminate variations in first-peak latency, the signals were truncated in such a way that a first major peak emerged at 2 ms. The total duration of the stimuli was 92 ms. They were gated on and off with a 2-ms cosine-squared ramp to minimize spectral splatter. Informal listening revealed that none of the tones produced an audible click at onset or offset.

The stimulus waveforms are displayed in the lower three rows of Fig. 3. The vertical arrows below each panel indicate the time point of 2 ms, to which the first major peak of a stimulus was aligned. For SIN tones, the waveform exhibits a clear period corresponding to F0. For ALT tones, intervening peaks produce a pseudo-period corresponding to 2F0, so the waveforms of the ALT tones look like octave-shifted SIN tones (indicated by diagonal arrows). The waveforms of the RAND tones are more irregular, provided that they are built up of many, poorly resolved components.

2.2. Simulated auditory representations

Several stages of the internal processing of the stimuli were simulated using a recent version of the Auditory Image Model (AIM), referred to as the *drnl/med/sti* model (Patterson and Irino, 1998). A compound bandpass filter simulated the outer and middle ear transfer characteristic; it is the sum of four parallel, 2nd order bandpass filters: 100–1300 Hz, gain of −12 dB; 350–6500 Hz, gain of 1.5 dB; 1800–5200 Hz, gain of 5 dB; 7500–14,000 Hz, gain of −11 dB. Basilar membrane motion (BMM) was simulated with a dual-resonance nonlinear filterbank (Lopez-Poveda and Meddis, 2001; Meddis et al., 2001); 70 frequency channels covered the frequency range between 500 and 10,000 Hz. The inner hair cell receptor potentials were simulated using the parameters recommended by Shamma et al. (1986). The spike probability at the synapse between the inner hair cell and the auditory nerve fibre were simulated using the specifications of Sumner et al. (2002); there were 500 high-spontaneous-rate fibers per channel. The output of this stage is a time-frequency representation of the firing pattern of the auditory nerve, often referred to as the 'neural activity pattern' (NAP) (Patterson et al., 1995). A 'summary NAP' is then produced by averaging across frequency channels; it represents the overall activity of the auditory nerve as a function of time.

The pitch extraction stage was simulated with AIM (Patterson et al., 1995) using strobed temporal integration (Patterson et al., 1992). This time-domain model of pitch processing assumes that the auditory system transforms the fragile spike-timing information in the NAP to a more stable time-interval representation. In the initial stage, interspike-interval histograms are computed separately for each frequency channel; the tonotopic array of interspike-interval histograms is the auditory image (AI). The histograms are then averaged across the array to produce a summary AI. For harmonic complexes, the summary AI shows a major peak
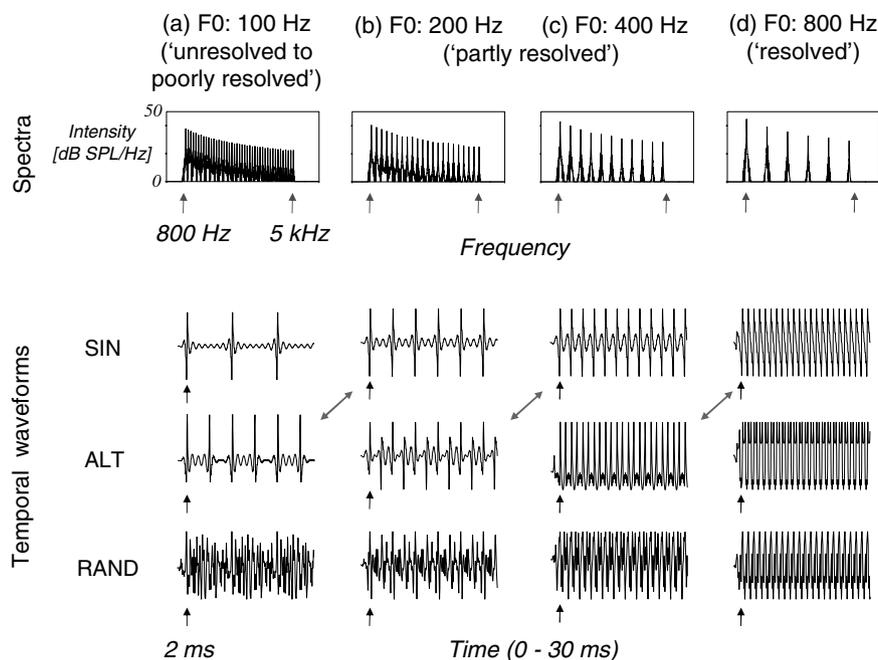

Fig. 3. Upper row: Amplitude spectra of the stimuli. The tones had F0s of 100, 200, 400, and 800 Hz, and they included all harmonics between 800 and 5000 Hz (marked by vertical arrows below each panel). (a) Missing F0 of 100 Hz; harmonics 8–50 ('transition from unresolved to poorly resolved'). (b) Missing F0 of 200 Hz; harmonics: 4–25 ('partly resolved'). (c) Missing F0 of 400 Hz; harmonics 2–12 ('partly resolved'). (d) F0 of 800 Hz; harmonics 1–6 ('resolved'). The amplitudes of the harmonics decreased by 6 dB per octave. Lower three rows: Stimulus waveforms (first 30 ms). SIN tones have a period corresponding to F0. ALT tones exhibit a pseudo-period corresponding to 2F0, so their waveforms look similar to those of octave-shifted SIN tones (indicated by diagonal arrows). The waveforms of the RAND tones are more irregular, provided that they are built up of many, poorly resolved components. All tones show a first major peak around 2 ms (arrows at the bottom of each panel).

at the time-interval associated with the period of the signal. Thus, the model predicts that listeners will hear a pitch at the fundamental frequency, regardless of whether this component is physically present or not (Meddis and Hewitt, 1991; Patterson et al., 1992).

Yost et al. (1996) showed that the height of the first peak in the autocorrelation function of a sound is closely related to the salience, or strength of the pitch of a sound. Shofner and Selas (2002) have recently shown that the pitch strength follows a modified version of Stevens' power law, $\Psi = k\Phi^{\beta}$ (Ψ: magnitude of the sensation, Φ: intensity of the stimulus, k: constant factor, β: exponent). The exponent β is typically greater than 1, indicating that pitch salience grows faster than the height of the first peak in the autocorrelation function. AIM is intended to summarize physiological activity associated with pitch extraction (Patterson et al., 1995), and so it does not include the non-linear growth of pitch strength described by Shofner and Selas (2002).

Cariani and Delgutte (1996a,b) related the pooled interspike-interval statistics of the auditory nerve of anesthetized cats to psychoacoustic results on the perceived pitch of various sound types, such as harmonic and inharmonic amplitude modulated tones, quasi-frequency-modulated tones, amplitude modulated noise, single-formant vowels, and click trains. They found a high correspondence between the perceived pitch and its salience and the all-order interspike interval distributions of the cat auditory nerve, thus providing physiological evidence for temporally-based pitch models.

2.3. MEG-measurements

2.3.1. Subjects

Seven subjects (2 male and 5 female) participated in the experiment. All participants were paid for their services. They had no history of otological or neurological disorders, they had normal audiological status, and they reported to be right-handers. After having explained the nature and purpose of this study, consent was obtained. The experimental procedures were conducted in accordance with the Ethics Commission of the University of Münster and the Declaration of Helsinki.

2.3.2. Neuromagnetic recordings

The stimuli were played out by a speaker (compressor driver type) installed outside the magnetically shielded chamber. They were fed through a 6.3-m plastic tubing to a silicon ear piece inside the subject's ear. The 19-ms delay induced by the transmission system was compensated for. The auditory evoked fields were recorded with a 37-channel first-order gradiometer system (Biomagnetic Technologies) in a magnetically shielded room. The subjects lay on their left side; body movements were minimized by a special vacuum cushion. The locations of the sensors relative to the head were determined by a sensor position indicator system (Polhemus 3 space tracker). The sensor array was positioned over the right hemisphere as close as possible to the head surface and manually centered above a point located about 1.5 cm superior to the location T4 of the 10–20 system for electrode placement. Subjects were asked to stay awake and to keep their eyes open. In order to maintain their attention and to reduce eye movements, a silent video was transmitted into the measurement chamber through optical fibers.

The stimulus intensity was set to 60 dB above subjective hearing threshold, as determined for the 100-Hz SIN stimulus. This threshold was also adopted for the other conditions, which were matched for RMS-level. The time interval between the onsets of successive stimuli (stimulus onset asynchrony; SOA) was randomized between 4 and 5 s (mean 4.5 s). The stimuli were presented to the left ear in randomized order in 6 blocks, each lasting about 10 min. The data were acquired in two experimental sessions. The number of epochs collected per session and stimulus condition was 66.

2.3.3. Data processing

Data (recording bandwidth 0.1–100 Hz) were sampled at 1041.7 Hz using 16-bit analog to digital converters and stored on hard disk for further analysis. Epochs with a maximum amplitude exceeding 2 pT were considered to be contaminated by artifacts, such as eye blinks, body movements, or external disturbances; such epochs were discarded. The remaining data were averaged separately for each stimulus condition. The resulting average was finally baseline corrected (time window beginning 300 ms prior to stimulus onset) and low-pass filtered (0–20 Hz).

The latencies and amplitudes of the N100m deflection were determined automatically by establishing the RMS-maximum of the magnetic activity registered by all 37 sensors during the relevant time range (70–120 ms after stimulus onset).

2.3.4. Statistical analysis

The amplitudes and latencies of the N100m deflection were analyzed in two separate, repeated-measures analyses of variance (ANOVAs) with 'F0', 'Relative phase', and 'Session' as the independent variables.

3. Results

3.1. Simulated stimulus transformations along the auditory pathway

3.1.1. Basilar membrane motion (BMM)

The tonotopic dimension along the basilar membrane is quasi-logarithmic, whereas the harmonics of a complex tone are linearly spaced in frequency. As a result, the low harmonics are more 'resolved' than the higher harmonics, and the higher harmonics interact on the basilar membrane, producing a modulation at the rate of F0 (Schouten, 1938). Since the frequency range of our stimuli was fixed (800–5000 Hz), the number of harmonics in the stimulus decreased and the resolution increased with F0. This is illustrated in Fig. 4, which shows the first 30 ms of the BMM for the two extreme F0-conditions: 100 Hz (left) and 800 Hz (right). Each line represents the activity of a cochlear frequency channel with an equivalent rectangular bandwidth (ERB) of one (Lopez-Poveda and Meddis, 2001); the center frequencies are displayed with a relative offset of 2.85 ERBs.

For the 100-Hz-stimulus with harmonics 8–50, the BMM is largely determined by the interaction of unresolved harmonics: in the SIN condition, this produces a 100-Hz beating shown by the dark vertical ridges separated by 10 ms; in the ALT condition, the pseudo-period gives the impression that the beat is at 200 Hz; in the RAND condition, the period can hardly be identified. Note, however, that there is less interaction at the lower edge of the spectrum, because the lowest harmonics are fairly resolved. In the following, the 100-Hz conditions will be referred to as being at the transition from poor to no spectral resolution.

The 200- and 400-Hz tones are not displayed in Fig. 4; they contain harmonics 4–25 and 2–12, respectively, and their spectra are partially resolved. For the 800-Hz tones (Fig. 4, right), six horizontal ridges show that harmonics 1–6 are resolved. As there is almost no interaction between these components, there is no phase-dependent temporal regularity, and so the BMMs look quite similar for the three phase conditions.

3.1.2. Temporal firing pattern of the auditory nerve (summary NAP)

Fig. 5 shows the summary NAPs, representing the overall activity of the auditory nerve. The activity is displayed
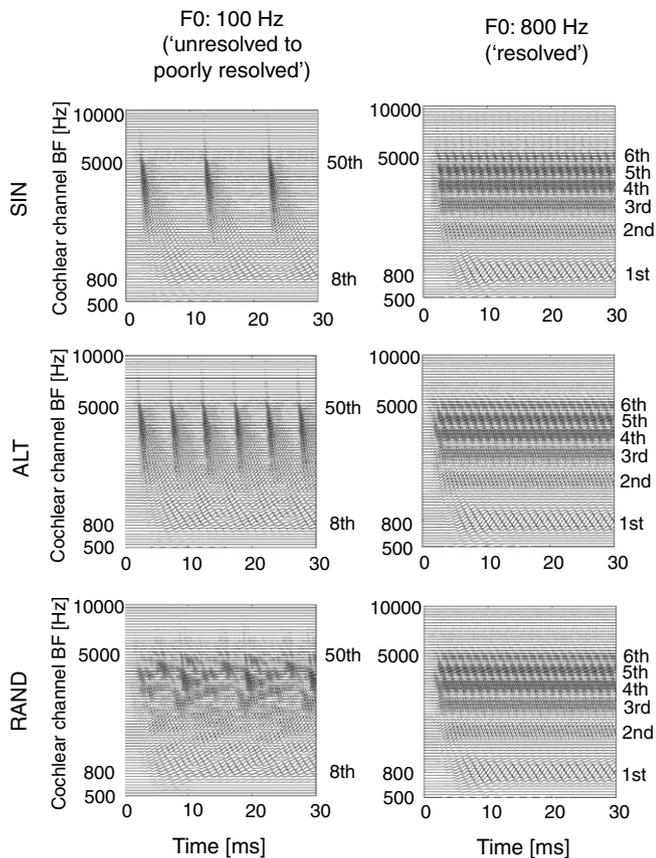
Fig. 4. Basilar membrane motion (BMM) during the first 30 ms after stimulus onset. The cochlear tonotopic axis, represented by channel best frequencies (BF), is indicated at the left margin of each panel. The harmonics included in the stimuli are indicated at the right margin. Left column: 100-Hz stimuli (harmonics 8–50), being at the transition from poor to no spectral resolution. The BMM is dominated by mechanical interaction, which depends on the relative phase of the stimulus components. For the 100-Hz SIN tone, there is beating at 100 Hz (dark vertical ridges at a relative distance of 10 ms); for the ALT tone, the beating is twice as fast; the pattern of the RAND tone is more irregular. Right column: Spectrally resolved 800-Hz stimuli (harmonics 1–6). The BMM is characterized by separated bands of activity, reflecting the increased frequency spacing of the components, and the concomitant lack of interaction.

for the first 30 ms after stimulus onset. In all conditions, the first major peak occurs after 3–4 ms (two vertical lines), showing that the effective onsets of the stimuli were well matched. For the poorly resolved tones with a low missing F0 (100 and 200 Hz), the stimulus waveform is readily recognizable from the summary NAP (compare with Fig. 2). With increasing spectral resolution the neural representations become more similar and progressively independent of the relative phase of the frequency components. The summary NAPs for poorly and partly resolved ALT tones are characterized by the pseudo-period, making them practically indistinguishable from the summary NAPs of SIN tones with twice the F0 (see diagonal arrows). The summary NAPs for RAND tones are more irregular and the period is hardly recognizable, especially if the number of frequency components is high (poor spectral resolution).

3.1.3. The pitch-extraction stage (summary AI)

The summary AI represents the neural interspike-interval statistics for a given stimulus, averaged across all frequency channels. In the case of harmonic tones, it is characterized by a peak at the period of the sound, F0. This is the model's correlate of the pitch we hear. The height of the peak (relative to the background activity) provides a measure of pitch salience (Wiegrebe et al., 2000; Shofner and Selas, 2002). Fig. 6 displays the summary AIs for all stimulus conditions. The dotted lines indicate prominent interspike-interval peaks at 10 ms (100 Hz), 5 ms (200 Hz), 2.5 ms (400 Hz), and 1.25 ms (800 Hz), respectively. With respect to the position of that peak, there is no difference between the SIN, ALT, and RAND tones. Moreover, the peaks have the same relative magnitudes for the 200, 400, and 800 Hz-stimuli, but not for the 100-stimulus. This suggests that all tones produce a pitch corresponding to F0, and that for the 100 Hz tone the pitch salience is highest in SIN phase and lowest in RAND phase. Both predictions were confirmed by informal listening.

The stability of the indicated pitch values, irrespective of relative phase, can be explained as follows: All stimuli, including the 100 Hz condition, contained at least a few spectral components in the lower frequency region, which were reasonably resolved (cf. Fig. 4). The resolved components consistently contributed to the F0-peak in the summary AI (Fig. 6). The unresolved components, stimulating higher frequency channels, contributed to the F0-peak in SIN tones and to both the F0- and the 2F0-peak in ALT tones; in RAND tones they produced a more irregular pattern. Summed up across all frequency channels, the F0-peak was dominant for all conditions, indicating that the relative phase of the components did not affect the prevailing pitch. It should be noted, however, that for stimuli with a spectral resolution even poorer than that of our 100 Hz-tone (lowest harmonics not resolved any longer), the ALT-condition would produce a 2F0-peak just as high as the F0-peak. As the first major peak in the summary AI, corresponding to the first-order interspike interval, is most important for the pitch estimate (Kaernbach and Demany, 1998), a change from SIN to ALT in completely unresolved tones induces an upward octave shift in pitch (Shackleton and Carlyon, 1994). In our study, the secondary peak at 5 ms, seen for the 100 Hz ALT tone (arrow in Fig. 6) is still smaller than the peak at 10 ms, suggesting that the pitch of the stimulus corresponds more strongly to the lower octave.

3.1.4. Simulation-based predictions

If the latency of the N100m response were largely determined by peripheral processes, as characterized by the summary NAP in our simulations, the N100m should be sensitive to the relative phase of the stimulus components when the F0 is 100 Hz, and the phase-sensitivity should decrease with F0 and be absent at 800 Hz. Moreover, the
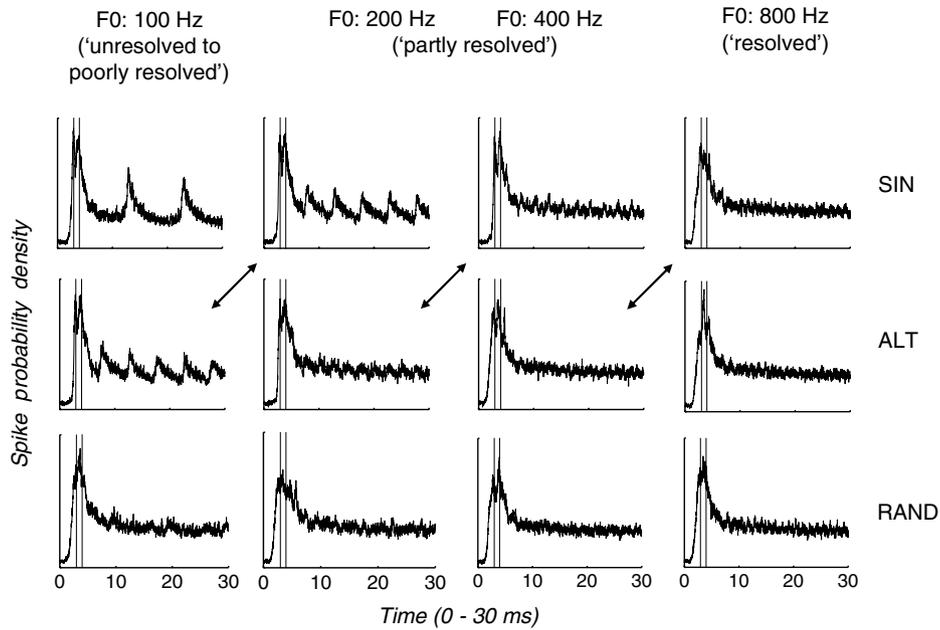
Fig. 5. Summary neural activity patterns (summary NAPs) of auditory nerve activity during the first 30 ms after stimulus onset. For all 12 conditions, the activity shows a first major peak after 3–4 ms (time interval marked by two vertical lines). Thus, the onset characteristics of the stimuli were well matched. Two effects can be noticed: First, the influence of the phase spectrum on the summary NAP decreases as the stimuli become spectrally resolved by the auditory periphery. Second, the pseudo-period of the unresolved and partly resolved ALT tones appears in the summary NAP, making it practically indistinguishable from the summary NAP of SIN tones with twice the F0 (diagonal arrows).
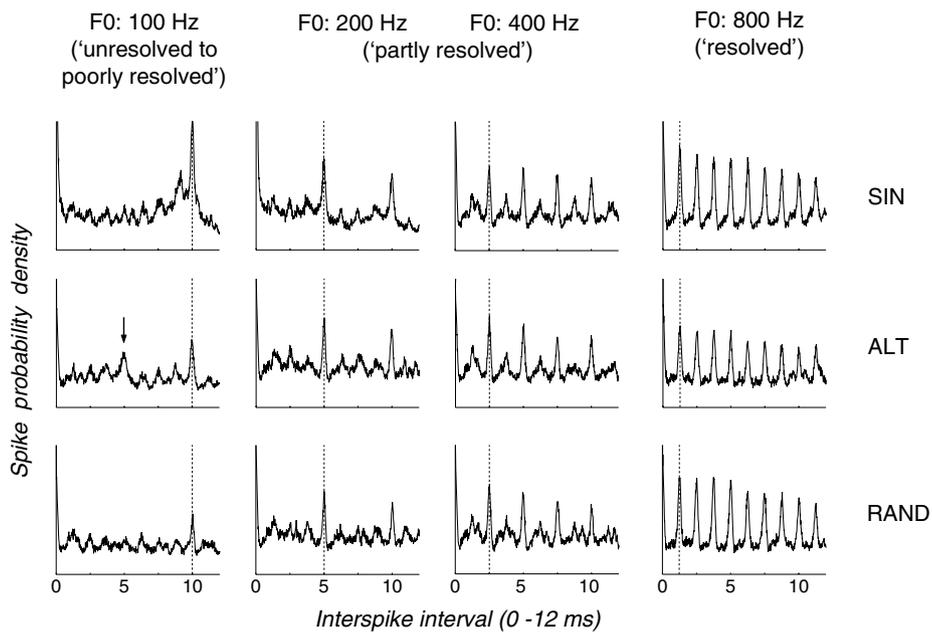

Fig. 6. Summary auditory images (summary AIs) representing the interspike-interval statistics of auditory nerve firing. The abscissa is the interspike time. The perceived pitch is represented by the first-order interspike interval (first major peak after trigger peak at time 0); the salience of the pitch corresponds to the relative height of that peak. The summary AIs predict a pitch corresponding to F0 (dotted lines) for all conditions.

latency of the N100m in response to ALT stimuli containing unresolved components should be similar to that for SIN stimuli with twice the F0. If, on the other hand, the N100m were related to more central pitch extraction, it should be sensitive to F0, but not to the phase-spectrum of the tones.

3.2. Neuromagnetic findings

Fig. 7 shows, for one representative subject, the magnetic fields evoked by SIN tones with different F0s. Signals are displayed for a time interval of 0–160 ms after stimulus onset. The overlaid curves on the left side represent the

activity recorded by the 37 MEG-sensors. The curves on the right side show the corresponding RMS-values, derived from all measurement channels. The N100m deflection starts to emerge at about 60 ms after stimulus onset and reaches its maximum after 90–110 ms.

The latencies of the RMS maxima (marked by dotted lines) decrease with increasing F0 (solid reference lines at 100 ms). A statistical summary for all stimulus conditions and all subjects is presented in Fig. 8. The group data are evidently consistent with the individual data shown in Fig. 7. The significance of the F0 effect on the N100m latency was verified by an ANOVA ($F(3,6) = 20.3$, $p < 0.0001$). Bonferroni-post hoc testing revealed that apart from the 100- vs. 200-Hz comparison and the 400- vs. 800-Hz comparison, all differences between conditions were statistically significant.

The best-fitting line in linear-log coordinates is

$$\text{latency} = 93.7 - 2.8 \cdot \log_2 (\text{F0}/100) \text{ ms}.$$

Thus, every doubling of F0, corresponding to an octave jump in pitch, was associated with a decrease in N100m latency of about 2.8 ms. The function appearing as the trendline in Fig. 8 accounts for 99% of the variance in the data. Neither 'Relative phase' ($F(2,6) = 0.53$, $p = 0.6$) nor 'Session' ($F(1,6) = 0.02$, $p = 0.87$) had an effect on the N100m latency. Moreover, there were no significant interactions.

The ANOVA on the N100m amplitudes revealed no systematic effects whatsoever. Neither 'F0', nor 'Relative phase', nor 'Session' had a significant effect (Table 1).
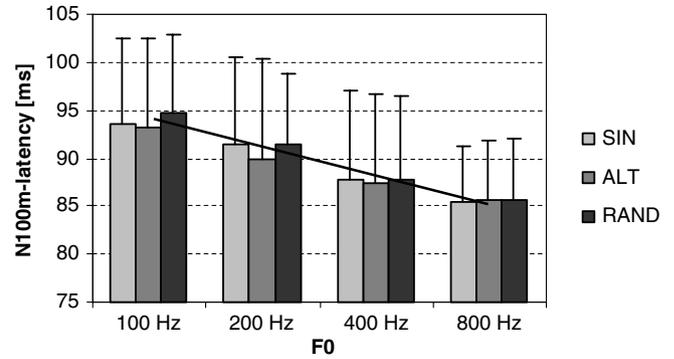

Fig. 8. Means and standard deviations of the N100m latency for the 12 stimulus conditions. Every doubling of F0, corresponding to an octave jump in pitch, was associated with a decrease in N100m latency of about 2.8 ms. Relative phase had no effect.

Table 1
ANOVA-results for N100m amplitude

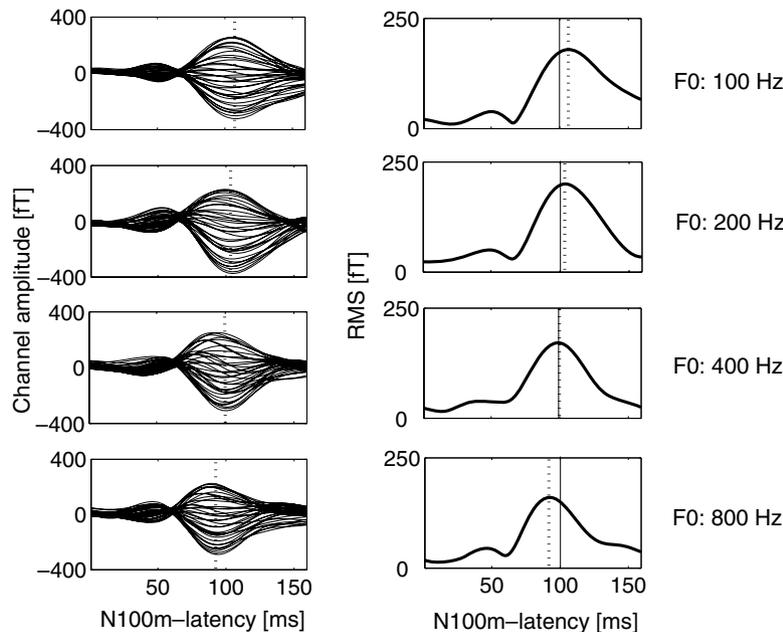|	Mean	F(df)	p
F0	100 Hz: 215 fT 200 Hz: 213 fT 400 Hz: 204.2 fT 800 Hz: 214.9 fT	$F(3,6) = 1.6$	0.22 n.s.
Relative phase	SIN: 213.3 fT ALT: 208.8 fT RAND: 213.4 fT	$F(2,6) = 3$	0.09 n.s.
Session	1: 213.2 fT 2: 210.4 fT	$F(1,6) = 0.2$	0.68 n.s.

Fig. 7. Auditory evoked fields (AEF) elicited by SIN tones with different periods (single subject). The plots on the left show an overlay of the signals recorded by the 37 measurement sensors. The root-mean-square value of these signals is shown on the right. The main peak around 100 ms (maximum indicated by dotted line) corresponds to wave N100m. Latency decreases steadily with increasing F0 (cf. solid reference line at 100 ms).

4. Discussion

A special set of harmonic complex tones was designed to clarify whether the AEF deflection N100m is essentially a function of the auditory periphery or of more central pitch-extraction processes. The latency of the N100m was found to decrease with increasing F0, while the relative phase of the stimulus components had no effect. In the following, we will compare this result with our model simulations, in which we considered not only the auditory periphery (basilar membrane motion, ongoing firing pattern of the auditory nerve), but also an interspike-interval representation, thought to underlie the pitch percept (Patterson et al., 1995).

4.1. Does the N100m reflect the auditory periphery?

The BMM (Fig. 4) and the summary NAP of the auditory nerve (Fig. 5) are highly sensitive to mechanical interaction between spectrally unresolved frequency components. Shape and regularity of the response patterns shown in these figures depend on the phase spectrum of the tones. Such effects are basically missing in the summary AI, thought to underlie the pitch percept (Fig. 6). In this respect, the latter model is consistent with our experimental data, which show that the latency of the N100m is not sensitive to the relative phase of the stimulus components, even if most of the harmonics are spectrally unresolved by the auditory periphery. This is incompatible with the view that the N100m latency is a direct function of the peripheral activation represented by the summary NAP.

4.2. Is the N100m a correlate of subjective pitch perception?

The auditory system is considered to require a certain number of stimulus cycles to reliably extract pitch (Forss et al., 1993). Consequently, the latency of AEF components affected by temporal pitch extraction should decrease with increasing stimulus periodicity (cf. Fig. 6). This is clearly the case for the N100m, suggesting that this component is somehow related to the pitch processing network. This does not necessarily imply that the N100m-generators are directly involved in temporal pitch extraction. According to the fMRI findings of Griffiths et al. (2001), the initial steps of temporal pitch extraction take place in the brainstem, the midbrain, and the thalamus, thus rendering it possible that the N100m is indirectly affected by that early processing. There is no reason to assume that the N100m reflects a latency code for pitch, as suggested by Crottaz-Herbette and Ragot (2000); this would require an internal temporal reference, which appears to be unavailable. We therefore consider the observed latency effect as a by-product of temporal pitch extraction.

The hypothesis that the N100m is a correlate of subjective pitch perception could only be true if the F0 were perceptually dominant in each of our stimuli. Shackleton and Carlyon, 1994 performed a psychophysical experiment, in which they investigated the pitch of SIN and ALT tones with varying degrees of spectral resolution. For completely unresolved tones (lower cutoff-frequency just below the 16th harmonic or higher), the pitch of SIN tones corresponded to F0, whereas the pitch of ALT tones was dominated by 2F0. For tones with slightly more spectral resolution (lower cutoff-frequency just below the 11th harmonic), the pitches of SIN and ALT tones did not substantially differ, and both were dominated by F0. The same stimulus set was used by Ragot and Crottaz (1998) in a MEG-study. For spectrally unresolved tones, the N100m latency was shorter for ALT phase than for SIN phase. Most of the stimuli that produced this effect had a poorer spectral resolution than our stimuli, suggesting that the pitches for the two phase conditions differed by one octave.

In our stimuli, the lowest component was the 8th (100-Hz stimulus), the 4th (200-Hz stimulus), the 2nd (400-Hz stimulus), and the 1st (800-Hz stimulus) harmonic, respectively; so no pitch differences were to be expected for SIN and ALT tones. This is consistent with our simulations, which showed that the first major peak in the summary AI is at the period corresponding to F0, for all our stimuli (Fig. 6). We therefore infer that the observed differences in N100m latency are related to pitch extraction. Our results are consistent with two experiments by Roberts et al. (1998). In the first experiment, pure tones of 1000 and 100 Hz were presented simultaneously at various amplitude ratios. At a certain ratio, the perceived pitch changed abruptly from high to low. This perceptual switch was associated with a categorical switch from short to long N100m-latencies. In the second experiment, a 1 kHz-tone amplitude-modulated with 100 Hz was used. The N100m latency did not correspond to the spectral components of the stimulus, but was significantly prolonged and corresponded to the latency for a 100 Hz pure tone. Thus, the N100m-latency did not reflect the objective spectral characteristics of the sound, but its perceptual attributes (missing F0 at 100 Hz).

4.3. Temporal stimulus processing along the auditory pathway

Woods et al. (1993) compared the latencies of wave V of the brainstem auditory evoked potential, of several middle-latency components, and of the late N100 in response to a 250 and a 4000 Hz pure tone. The latency was always shorter for the 4000-Hz tone, and the latency difference systematically increased with the absolute component latency. Similarly, Greenberg et al. (1998) noticed that the slope of the pitch-latency function is steeper for the N100 as compared to the P200, emerging about 100 ms later (data from Ragot and Lepaul-Ercole, 1996). They suggested that the 'latency expansion' along the pitch processing hierarchy reflects subsequent stages of temporal feature extraction. This view is supported by several neuro-imaging studies, which revealed a hierarchy of temporal pitch processing

along the auditory pathway, from the cochlear nucleus, the inferior colliculus, the medial geniculate body (Griffiths et al., 2001), to a non-primary cortical area in the lateral part of HG (Griffiths et al., 1998, 2001).

Several studies investigated the sensitivity of the N100m to both the onset- and the long-term characteristics of a sound (Joutseniemi et al., 1989; Forss et al., 1993; Alain et al., 1997; Biermann and Heil, 2000). Biermann and Heil (2000) gradually reduced the stimulus rise time from 128 to 2 ms and found a systematic increase in the N100m amplitude, associated with a decrease in latency. This led to the conclusion that the N100m is essentially a function of the onset characteristics of a sound, which are encoded already by the auditory periphery (Heil and Irvine, 1997). Alain et al. (1997) showed that, with increasing stimulus duration, the N100 systematically increased in amplitude and decreased in latency. They gave evidence that the N100 consists of subcomponents with different integration functions. In general, the integration times decreased with increasing frequency. Moreover, they suggested that the N100m-generators comprise both a short-time-constant system, sensitive to stimulus onset, and a long-time-constant system, which integrates neural activity up to about 60 ms. The time window of the latter system would be sufficient for an accurate periodicity analysis of most relevant sounds.

4.4. Alternative hypotheses

It is not possible to keep all spectral parameters of a harmonic complex tone (bandwidth and density of frequency components) constant while varying its period. In the present study, the number of harmonics decreased with increasing F0, while the spectral resolution increased (upper panels of Fig. 3). We cannot completely exclude that these spectral variations might have affected the N100m. There might have been a direct effect of spectral complexity or an indirect effect, via the concomitant variation in perceived pitch salience. In the following, we will discuss both possibilities.

4.4.1. Spectral complexity

There is vast evidence from the MEG-literature that the latencies of different AEF components decrease with increasing pitch. This was not only shown for pure tones (Rapin et al., 1966; Jacobson et al., 1992; Woods et al., 1993; Verkindt et al., 1995; Roberts and Poeppel, 1996; Stufflebeam et al., 1998; Lütkenhöner et al., 2001; Seither-Preisler et al., 2003), but also for synthesized speech sounds (Poeppel and Roberts, 1996), triangle and square waves (Roberts et al., 1998), bandpass-filtered harmonic tones (Ragot and Lepaul-Ercole, 1996; Ragot and Crottaz, 1998; Crottaz-Herbette and Ragot, 2000), and bandpass-filtered regular interval sounds (Krumbholz et al., 2003). In spite of substantial differences regarding the employed stimuli, the observed latency effects were basically consistent, suggesting that the effects were related to periodicity.

Two types of stimuli will be considered in more detail, now, a pure tone and a regular interval sound.

Pure tones represent only one frequency, which implies that the spectral complexity is independent of pitch. It has to be taken into account, however, that the travelling-wave delay on the basilar membrane decreases with frequency, which in turn might affect the measured AEF latencies. To compensate for this distortion, the travelling wave delays have to be subtracted from the respective N100m-latencies. The following example shows that, even after such a correction, the N100m-latency considerably decreases with frequency. For tones of 100, 200, and 400 Hz, Seither-Preisler et al. (2003) observed N100m latencies of 109.7, 101.2, and 91.6 ms. After compensation for the travelling wave delays according to Patterson (1994), the latencies were 96.5, 91.1, and 84.7 ms, respectively. This result cannot be explained in terms of the spectral-complexity hypothesis.

Regular interval sounds uniformly stimulate a cochlear region, just like noise, but are distinguished by a certain temporal regularity which is encoded in the interspike-interval statistics of the auditory nerve (Griffiths et al., 2001). Such stimuli produce a definite pitch, which demonstrates that pitch information is temporally encoded in the auditory system. In an earlier study (Krumbholz et al., 2003) we have identified an AEF component that specifically responds to the transition from a noise to a regular interval sound having the same long-term spectrum. The latency and amplitude of this 'pitch onset response' (POR) depended on pitch height and pitch salience, suggesting that the underlying generator plays a pivotal role in the formation of the final percept. Since the internal spectral representations of the stimuli were the same, the findings of Krumbholz et al. (2003) cannot be explained in terms of spectral processing. Thus, the latency of this AEF component appears to be determined mainly by the time required for temporal pitch extraction. In a subsequent study, Seither-Preisler et al. (2004) showed that the amplitude of the POR increases with the time elapsed since the preceding N100m. This interaction suggests that at least one of the N100m subcomponents is related to the pitch-sensitive POR generator.

4.4.2. Pitch salience

The pitch salience of a complex tone varies with its spectral composition. A tone consisting of many spectrally resolved components of low harmonic order produces a more salient pitch sensation than a tone consisting of few spectrally unresolved components of high harmonic order (Preisler, 1993). In the present study, the number of harmonics decreased with increasing F0, while the spectral resolution increased. These two circumstances have opposite effects on the perceived pitch salience.

In a recent fMRI-experiment, Penagos et al. (2004) showed that tones with high pitch salience elicit a stronger cortical activation than tones with low pitch salience. It may be speculated that the high cortical activation induced
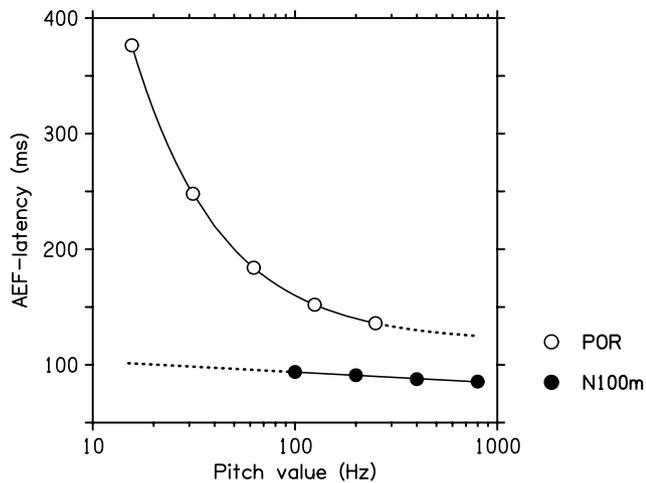
Fig. 9. Comparison of the pitch-latency functions for the POR (Krumbholz et al., 2003) and the N100m (present study).

by a salient stimulus gives rise to a relatively shorter AEF-latency. Two previous MEG-studies confirm the first part of the hypothesis, namely that the amplitudes of the N100m (Seither-Preisler et al., 2003) and the POR (Krumbholz et al., 2003) increase with pitch salience. However, the respective latencies were largely constant contradicting the idea that pitch-salience induced a higher cortical activation which translated into shorter AEF-latencies. Apart from that, the N100m-amplitudes were basically constant in the present study. This renders it unlikely that the observed latency effect arises from variations in pitch salience.

4.5. N100m versus POR

Our results suggest that the N100m is affected by processes related to temporal pitch extraction. This raises the question as to the functional significance of the N100m and the POR in the larger pitch processing network. The source of the POR is located near the medial part of Heschl's gyrus, somewhat anterior and inferior to the source of the N100m (Krumbholz et al., 2003). The pitch-latency function fitted to the MEG data of Krumbholz et al. (2003) is shown by the open circles in Fig. 9. The corresponding function from the current experiment is shown by the filled circles. The latency of the POR is substantially longer than that of the N100m. Although the two functions overlap only between 100 and 250 Hz, the extrapolated functions suggest a latency expansion between the N100m and the POR. This renders it unlikely that the two components reflect the same stage in the pitch processing hierarchy. Some of the N100m generators are presumably linked to the extended pitch processing network, although they do not constitute its final stage.

Acknowledgments

References

Alain, C., Woods, D.L., Covarrubias, D., 1997. Activation of duration-sensitive auditory cortical fields in humans. Electroencephalogr. Clin. Neurophysiol. 104, 531–539.

Beagley, H.A., Knight, J.J., 1967. Changes in auditory evoked response with intensity. J. Laryngol. Otol. 81, 861–873.

Biermann, S., Heil, P., 2000. Parallels between timing of onset responses of single neurons in cat and of evoked magnetic fields in human auditory cortex. J. Neurophysiol. 84, 2426–2439.

Cariani, P.A., Delgutte, B., 1996a. Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. J. Neurophysiol. 76, 1698–1716.

Cariani, P.A., Delgutte, B., 1996b. Neural correlates of the pitch of complex tones. II. Pitch shift, pitch ambiguity, phase invariance, pitch circularity, rate pitch, and the dominance region for pitch. J. Neurophysiol. 76, 1717–1734.

Cheung, S.W., Bedenbaugh, P.H., Nagarajan, S.S., Schreiner, C.E., 2001. Functional organization of squirrel monkey primary auditory cortex: responses to pure tones. J. Neurophysiol. 85, 1732–1749.

Crottaz-Herbette, S., Ragot, R., 2000. Perception of complex sounds: N1 latency codes pitch and topography codes spectra. Clin. Neurophysiol. 111, 1759–1766.

Eggermont, J.J., Ponton, C.W., 2002. The neurophysiology of auditory perception: from single units to evoked potentials. Audiol. NeuroOtol. 7, 71–99.

Forss, N., Makela, J.P., McEvoy, L., Hari, R., Sidtis, J.J., Volpe, B.T., 1993. Temporal integration and oscillatory responses of the human auditory cortex revealed by evoked magnetic fields to click trains. Hear. Res. 68, 89–96.

Greenberg, S., Poeppel, D., Roberts, T., 1998. A space-time theory of pitch and timbre based on cortical expansion of the cochlear travelling wave delay. Psychophys. Physiol. Adv. Hear., 293–300.

Griffiths, T.D., Buchel, C., Frackowiak, R.S., Patterson, R.D., 1998. Analysis of temporal structure in sound by the human brain. Nat. Neurosci. 1, 422–427.

Griffiths, T.D., Uppenkamp, S., Johnsrude, I., Josephs, O., Patterson, R.D., 2001. Encoding of the temporal regularity of sound in the human brainstem. Nat. Neurosci. 4, 633–637.

Heil, P., Irvine, D.R.F., 1997. First-spike timing of auditory-nerve fibers and comparison with auditory cortex. J. Neurophysiol. 78, 2438–2454.

Jacobson, G.P., Lombardi, D.M., Gibbens, N.D., Ahmad, B.K., Newman, C.W., 1992. The effects of stimulus frequency and recording site on the amplitude and latency of multichannel cortical auditory evoked potential (CAEP) component N1. Ear Hear. 13, 300–306.

Joutseniemi, S.L., Hari, R., Vilkman, V., 1989. Cerebral magnetic responses to noise bursts and pauses of different durations. Audiology 1989, 325–333.

Kaernbach, C., Demany, L., 1998. Psychophysical evidence against the autocorrelation theory of auditory temporal processing. J. Acoust. Soc. Am. 104, 2298–2306.

Krumbholz, K., Patterson, R.D., Seither-Preisler, A., Lammertmann, C., Lütkenhöner, B., 2003. Neuromagnetic evidence for a pitch processing center in Heschl's gyrus. Cereb. Cortex 13, 765–772.

Lopez-Poveda, E.A., Meddis, R., 2001. A human nonlinear cochlear filterbank. J. Acoust. Soc. Am. 110, 3107–3118.

Loveless, N., Levanen, S., Jousmaki, V., Sams, M., Hari, R., 1996. Temporal integration in auditory sensory memory: neuromagnetic evidence. Electroencephalogr. Clin. Neurophysiol. 100, 220–228.

Lütkenhöner, B., Lammertmann, C., Knecht, S., 2001. Latency of auditory evoked field deflection N100m ruled by pitch or spectrum? Audiol. NeuroOtol. 6, 263–278.

Lütkenhöner, B., Steinsträter, O., 1998. High-precision neuromagnetic study of the functional organization of the human auditory cortex. Audiol. NeuroOtol. 3, 191–213.

Meddis, R., Hewitt, J., 1991. Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I. Pitch identification. J. Acoust. Soc. Am. 89, 2866–2882.

Meddis, R., O'Mard, L.P., Lopez-Poveda, E.A., 2001. A computational algorithm for computing nonlinear auditory frequency selectivity. J. Acoust. Soc. Am. 109, 2852–2861.

Näätänen, R., Picton, T., 1987. The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. Psychophysiology 24, 375–425.

Patterson, R.D., 1994. The sound of a sinusoid: spectral models. J. Acoust. Soc. Am. 96, 1409–1418.

Patterson, R.D., Allerhand, M.H., Giguere, C., 1995. Time-domain modelling of peripheral auditory processing: a modular architecture and a software platform. J. Acoust. Soc. Am. 98, 1890–1894.

Patterson, R.D., Irino, T., 1998. Modeling temporal asymmetry in the auditory system. J. Acoust. Soc. Am. 104, 2967–2979.

Patterson, R.D., Robinson, K., Holdsworth, J., McKeown, D., Zhang, C., Allerhand, M., 1992. Complex sounds and auditory images. In: Cazals, Y., Demany, L., Horner, K. (Eds.), Auditory Physiology and Perception. Pergamon, Oxford, pp. 429–446.

Penagos, H., Melcher, J.R., Oxenham, A.J., 2004. A neural representation of pitch salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging. J. Neurosci. 24, 6810–6815.

Poeppel, D., Roberts, T.P.L., 1996. Effects of vowel pitch and task demands on latency and amplitude of the auditory evoked M100. Brain Cogn. 32, 156–158.

Preisler, A., 1993. The influence of spectral composition of complex tones and of musical experience on the perceptibility of virtual pitch. Percept. Psychophys. 54, 589–603.

Ragot, R., Crottaz, S., 1998. A dual mechanism for sound pitch perception: new evidence from brain electrophysiology. Neuroreport 9, 3123–3127.

Ragot, R., Lepaul-Ercole, R., 1996. Brain potentials as objective indexes of auditory pitch extraction from harmonics. Neuroreport 7, 905–909.

Rapin, I., Schimmel, H., Tourk, L.M., Krasnegor, N.A., Pollak, C., 1966. Evoked responses to clicks and tones of varying intensity in waking adults. Electroencephalogr. Clin. Neurophysiol. 21, 335.

Roberts, T., Ferrari, P., Poeppel, D., 1998. Latency of evoked neuromagnetic M100 reflects perceptual and acoustic stimulus attributes. Neuroreport 9, 3256–3269.

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

Roberts, T.P.L., Poeppel, D., 1996. Latency of auditory evoked M100 as a function of tone frequency. Neuroreport 7, 1138–1140.

Sams, M., Hari, R., Rif, J., Knuutila, J., 1993. The human auditory sensory memory trace persists about 10 s: neuromagnetic evidence. J. Cogn. Neurosci. 5, 363–370.

Schouten, J.F., 1938. The perception of subjective tones. Proc. Ned. Akad. Wet. 41, 1086–1094.

Seither-Preisler, A., Krumbholz, K., Lütkenhöner, B., 2003. Sensitivity of the neuromagnetic N100m deflection to spectral bandwidth: a function of the auditory periphery? Audiol. NeuroOtol. 8, 322–337.

Seither-Preisler, A., Krumbholz, K., Patterson, R., Seither, S., Lütkenhöner, B., 2004. Interaction between the neuromagnetic responses to sound energy onset and pitch onset suggests common generators. Eur. J. Neurosci. 19, 3073–3080.

Shackleton, T.M., Carlyon, R.P., 1994. The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination. J. Acoust. Soc. Am. 95, 3529–3540.

Shamma, S.A., Chadwick, R.S., Wilbur, W.J., Morrish, K.A., Rinzel, J., 1986. A biophysical model of cochlear processing: intensity dependence of pure tone responses. J. Acoust. Soc. Am. 80, 133–145.

Shofner, W.P., Selas, G., 2002. Pitch strength and Stevens's power law. Percept. Psychophys. 64, 437–450.

Stufflebeam, S.M., Poeppel, D., Rowley, H.A., Roberts, T.P.L., 1998. Peri-threshold encoding of stimulus frequency and intensity in the M100 latency. Neuroreport 9, 91–94.

Sumner, C., Lopez-Poveda, E.A., O'Mard, L.P., Meddis, R., 2002. A revised model of the inner-hair cell and auditory-nerve complex. J. Acoust. Soc. Am. 111, 2178–2188.

Verkindt, C., Bertrand, O., Perrin, F., Echallier, J.-F., Pernier, J., 1995. Tonotopic organization of the human auditory cortex: N100 topography and multiple dipole model analysis. Electroencephalogr. Clin. Neurophysiol. 96, 143–156.

Wiegrebe, L., Hirsch, H.S., Patterson, R.D., Fastl, H., 2000. Temporal dynamics of pitch strength in regular-interval noises: effect of listening region and an auditory model. J. Acoust. Soc. Am. 107, 3343–3350.

Woods, D., Alain, C., Covarrubias, D., Zaidel, O., 1993. Frequency-related differences in the speed of human auditory processing. Hear. Res. 66, 46–52.

Yost, W.A., Patterson, R., Sheft, S., 1996. A time domain description for the pitch strength of iterated rippled noise. J. Acoust. Soc. Am. 99, 1066–1078.