Duifhuis pitch: neuromagnetic representation and auditory modeling

Martin Andermann,1,2 Roy D. Patterson,3 Michael Geldhauser,1 Norman Sieroka,4 and André Rupp1

1Section of Biomagnetism, Department of Neurology, University Hospital of Heidelberg, Heidelberg, Germany; 2Section of Experimental Psychopathology, Department of Psychiatry, University Hospital of Heidelberg, Heidelberg, Germany; 3Centre for the Neural Basis of Hearing, Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge, United Kingdom; and 4Swiss Federal Institute of Technology Zurich, Zurich, Switzerland

Submitted 19 December 2013; accepted in final form 18 August 2014

Andermann M, Patterson RD, Geldhauser M, Sieroka N, Rupp A. Duifhuis pitch: neuromagnetic representation and auditory modeling. J Neurophysiol 112: 2616–2627, 2014. First published August 20, 2014; doi:10.1152/jn.00898.2013.—When a high harmonic is removed from a cosine-phase harmonic complex, we hear a sine tone pop out of the perception; the sine tone has the pitch of the high harmonic, while the tone complex has the pitch of its fundamental frequency, \( f_0 \). This phenomenon is commonly referred to as Duifhuis Pitch (DP). This paper describes, for the first time, the cortical representation of DP observed with magnetoencephalography. In experiment 1, conditions that produce the perception of a DP were observed to elicit a classic onset response in auditory cortex (P1m, N1m, P2m), and an increment in the sustained field (SF) established in response to the tone complex. Experiment 2 examined the effect of the phase spectrum of the complex tone on the DP activity: Schroeder-phase negative waves elicited a transient DP complex with a similar shape to that observed with cosine-phase waves but with much longer latencies. Following the transient DP activity, the responses of the negative and positive Schroeder-phase waves converged, and the increment in the SF slowly died away. In the absence of DP, the two Schroeder-phase conditions with low peak factors both produced larger SFs than cosine-phase waves with large peak factors. A model of the auditory periphery that includes coupling between adjacent frequency channels is used to explain the early neuromagnetic activity observed in auditory cortex.

Duifhuis pitch: phase; auditory cortex; magnetoencephalography; auditory modeling

MUSICAL TONES, THE VOCALS OF SPEECH, AND MANY ANIMAL CALLS ARE QUASI-PERIODIC SOUNDS WITH PEAKY WAVEFORMS. THEIR MAGNITUDE SPECTRA ARE COMPOSED OF SETS OF ADJACENT HARMONICS OF A FUNDAMENTAL WHOSE FREQUENCY, \( f_0 \), IS THE REPETITION RATE OF THE WAVE, AND THE DOMINANT HARMONICS ARE MODE LOCKED (FLETCHER AND ROSSING 1998), THAT IS, THEIR RELATIVE PHASES BEAR A SIMPLE, FIXED RELATIONSHIP TO THE PHASE OF THE FUNDAMENTAL. THE HARMONICS FUSE IN PERCEPTION TO PRODUCE A MUSICAL TONE WITH A DISTINCTIVE TIMBRE AND A PROMINENT PITCH CORRESPONDING TO \( f_0 \), EVEN WHEN THERE IS NO ENERGY AT \( f_0 \) (SCHOUTEN ET AL. 1962). MAGNETOENCEPHALOGRAPHY (MEG) SHOWS THAT, WHEN THESE HARMONIC COMPLEX TONES (HCTs) WITH SMOOTH PHASE SPECTRA ARE PRESENTED TO LISTENERS, THEY PRODUCE A STRONG RESPONSE WITH A CHARACTERISTIC TRIPARTITE DEFLECTION AT STIMULUS ONSET (P1m, N1m, P2m), AND A LARGE SUSTAINED FIELD (SF) THAT BUILDS UP OVER ABOUT 400 MS AND CONTINUES FOR THE DURATION OF THE TONE (E.G., GUTSCHALK ET AL. 2004). IN THE AUDITORY LITERATURE, THERE ARE MANY DIFFERENT NAMES FOR THE PITCH PERCEPTIONS PRODUCED BY HCTs (SEE HARTMANN 1996 OR PRESSNITZER ET AL. 2001). HERE, WE ADOPT THE TERM “MUSICAL PITCH” (MP) TO DISTINGUISH THIS FORM OF PITCH FROM “DUIFHUIS PITCH,” WHICH IS THE PRIMARY TOPIC OF THIS PAPER. THE TERM MUSICAL PITCH WAS PROPOSED BY GUTTMAN AND PRUZANSKY (1962) WHO THOUGHT IT IMPORTANT TO DISTINGUISH BETWEEN ANSI (1994) STANDARD PITCH, WHICH IS SIMPLY THE SENSATION THAT ALLOWS ONE TO ORDER NOTES ON A SCALE FROM LOW TO HIGH, AND MP, WHICH ALLOWS ONE TO RELIABLY CONVEY THE DIATONIC INFORMATION OF MUSIC, I.E., SEQUENCES OF MUSICAL INTERVALS.

Duifhuis pitch (DP) was originally demonstrated using a click train, which is perhaps the simplest HCT. It is composed of all harmonics of a given \( f_0 \) with all components in cosine phase, and it produces a very strong MP at \( f_0 \) (SHOFNER AND SELAS 2002). Duifhuis (1970, 1971) reported that, when one high harmonic of a click train is removed, a second pitch (DP) emerges in the perception, its frequency is close to that of the excluded harmonic, and its timbre is like that of a sinusoid rather than a musical tone. The experiments showed that DP is audible for a range of high harmonics when \( f_0 \) is low (25 to 100 Hz) and the period of the wave is relatively long (40 to 10 ms, respectively). The first experiment of the present paper was intended to determine whether MEG could be used to identify a physiological representation of DP in auditory cortex and distinguish it from the response to MP, in terms of both transient deflections and the SF.

Duifhuis (1970, 1971) argued that the feature that leads to the perception of a DP originates in the periphery (i.e., the cochlea); the panels of row 1 of Fig. 1 illustrate the argument. The stimuli in this and the other examples in this Introduction are a subset of those presented in the comprehensive paper on DP by Lin and Hartmann (1997). In this first example, the sound is a sine-phase (SIN) HCT, which produces a pulsive wave (row 1, left column, top panel). It is very similar to the click train produced by a cosine-phase (COS) HCT, but the SIN HCT has a bipolar pulse rather than a unipolar pulse, once per period of the sound. The \( f_0 \) in this example is 50 Hz, so the period is 20 ms. The same wave with the 19th harmonic removed is shown in the top panel of the right-hand column of the row; a low-level oscillation at 950 Hz (the 19th harmonic) can be seen running between the bipolar pulses in this wave. The bottom panels of row 1 show the “cochlear responses” (basilar membrane motion; BMM) that these sounds are thought to produce. In both of these cochlear responses, the bipolar pulses of the SIN HCT produce an array of filter impulse responses in the higher frequency channels, responses which die away rapidly, leaving a “quiet” interval within the period of the HCT. In this quiet interval, when the 19th
harmonic is canceled (right column), the basilar membrane vibrates at the frequency of the canceled harmonic in a band of channels at, and above, 950 Hz, despite the fact that there is no energy in the sound at the 19th harmonic. Duifhuis argued that it is this oscillating feature, generated peripherally, that leads to the sinusoidal pitch that is heard when a harmonic is removed. The theory also explains why we do not hear a DP for low harmonics of $f_0$; the impulse responses in the region of the low harmonics are long relative to the period of the HCT, and so there is no quiet interval in which to isolate the feature associated with a canceled harmonic.

Duifhuis (1972) tested his “peripheral origin” hypothesis by randomizing the starting phases of the harmonics of his HCT. A random-phase wave (RAND), and the same wave with the 19th harmonic removed, are shown in the left and right columns of row 2 of Fig. 1 (RAND; top panels). Note that it is no longer possible to discern the absence of the 19th harmonic in the wave in the right-hand column because of the complexity of the waveform. Similarly, in the cochlear response, randomizing relative phase disrupts the tidy pattern of impulse responses produced by the SIN HCT. Duifhuis (1972) correctly predicted that removing a high harmonic of a RAND HCT would not induce a DP because there was no quiet interval within the period of the sound and no sinusoidal feature within that quiet interval.

Lin and Hartmann (1997) replicated Duifhuis’ early experiments using both SIN and RAND waves. They also extended the range of harmonics shown to produce DP. They illustrated Duifhuis’ hypothesis about the peripheral origin of the phenomenon using a traditional, gammatone auditory filterbank (GT-AFB; e.g., Patterson 1994) to simulate the cochlear responses to their stimuli (Lin and Hartmann 1997; see their Figs. 6–8). Their cochlear responses are similar to those shown in the SIN and RAND rows of Fig. 1 in most respects. There is, however, one important difference, which is that the sinusoidal feature associated with DP is restricted to a very narrow band of channels in the output of the GT-AFB centered on the frequency of the excluded harmonic. We return to this difference in the DISCUSSION, where we argue that it might explain Lin and Hartmann’s observation that the emergence of a DP requires a quiet interval that extends across a surprisingly broad range of frequencies.

The results of the first MEG experiment in the present paper show, as expected, that there is a prominent transient response in auditory cortex when a harmonic is removed from a SIN HCT and we hear a DP. Moreover, there is no transient MEG response when a harmonic is removed and we do not hear a DP (either because the frequency of the excluded harmonic is too low or the relative phases of the components have been
randomized). The details are presented in the results section for the first experiment.

The strength of the DP feature in the MEG data led to a second MEG experiment, again prompted by the research reported in Lin and Hartmann (1997). In this case, the MP stimulus was a Schroeder-phase (SCHR) wave (Schroeder 1970) in which the phase of the $n$th harmonic, $\phi_n$, is progressively shifted to produce the phase spectrum

$$\phi_n = (+/-) \pi n(n - 1)/N \text{ (radians)} \quad (1)$$

where $N$ is the total number of harmonics. These SCHR waves have a low peak factor, like RAND waves, but the instantaneous frequency increases (SCHR−) or decreases (SCHR+) in an orderly way during the period of the sound, depending on the sign of the phase term. A typical SCHR− wave is shown in the top panel of row 3 of Fig. 1, left column; the wave for the same sound with the 19th harmonic removed is shown in the right column. The removal of the high harmonic is not obvious in the wave. Nevertheless, Lin and Hartmann (1997) showed that these SCHR− sounds induce a DP when a high harmonic is removed from the sound. They explained the phenomenon using cochlear responses like those below the SCHR− waves in row 3 (see Fig. 8 in Lin and Hartmann 1997). Specifically, the progressive phase shift alters the temporal alignment of the “impulse responses” of the SIN wave, (row 1 of Fig. 1), but there remains a small, quiet interval at some point within the period of the cochlear response for all high-frequency channels. As a result, when the 19th harmonic is removed, the vibration in the quiet interval is audible (and just visible in the right column of row 3 in channels at, and above, 0.95 kHz). Thus the results of Lin and Hartmann support the peripheral origin hypothesis concerning DP.

In the second experiment of the present paper, we measure the MEG response in auditory cortex with SCHR− waves and confirm that there is a significant deflection shortly after the removal of the 19th harmonic when the listener hears a DP. We also measure the MEG response with SCHR+ waves because the cochlear response to SCHR+ waves is not simply a time-reversed version of the response to the SCHR− wave. This is because the stimulus phase spectrum interacts with the phase characteristic of the basilar membrane (e.g., Kohlrausch and Sander 1995). The waves and cochlear responses for the SCHR+ waves are shown in row 4 of Fig. 1. The progressive, positive phase shift alters the temporal alignment of the channels and the depth of modulation over the course of the period, in a way that increases the size of the quiet interval within the period of the cochlear response. As a result, the effect of removing the 19th harmonic (right column) is clearly visible. The BMM plots suggest that the DP might be more audible in SCHR+ waves than in SCHR− waves, and we might expect to find that the DP response in the auditory evoked field (AEF) differs for the SCHR+ and SCHR− waves. The fact that the depth of modulation in the cochlear response differs between SCHR+ and SCHR− waves is well known (e.g., Carlyon and Datta 1997a, 1997b; Dau et al. 2000; Kohlrausch and Sander 1995; Rupp et al. 2008; Uppenkamp et al. 2001).

### EXPERIMENT 1

#### MATERIALS AND METHODS

**Participants**

Thirteen adult listeners (six women; one left-handed; mean age: 29.9 ± 10.7 yr) volunteered for experiment 1. The subjects provided written, informed consent prior to the experiment. None of the participants reported any history of peripheral or central hearing impairment or any psychiatric or neurological disorder. Normal hearing was confirmed using a brief audiometric screening procedure. The experiments in this paper were approved by the local ethics committee (Medical Faculty, University of Heidelberg).

**Stimuli**

All stimuli were generated using MATLAB 7.1 (The Mathworks). The basic stimulus was an HCT with a $f_0$ of 50 Hz and 45 equal-amplitude harmonics. The highest two harmonics were progressively attenuated in amplitude to reduce Gibbs phenomena and the perception of an edge pitch (amplitude factors: $A_{N-1} = 0.6; A_N = 0.3$).

In an effort to direct listener’s attention to any DP, we employed a pop-out paradigm in which each stimulus consisted of two 800-ms-long segments: in the first segment, all harmonic components were present within the stimulus, and MP was clearly audible; then, in the second segment, a single harmonic was abruptly removed to draw attention to the DP. This paradigm also facilitates the extraction of the neuromagnetic response associated with the removal of a single harmonic, separate from the cortical activity associated with energy onset (Biermann and Heil 2000; Gutschalk et al. 2002) and the perception of MP at the onset of the HCT. In the SIN2 and RAND2 conditions, the second harmonic (100 Hz) was removed at the transition to the second segment, and in the SIN19 and RAND19 conditions, the 19th harmonic (0.95 kHz) was removed at the transition to the second segment. In the SIN2 and SIN19 conditions, all harmonics began in sine phase; in the RAND2 and RAND19 conditions, the harmonics began in random phase. The specific phase configuration was chosen (by informal listening) to minimize the weak, diffuse change percept at the transition from the first to the second segment of the RAND stimulus. The peripheral origin hypothesis predicts that DP will only be heard in the SIN19 condition where the omitted harmonic elicits detectable activity in the corresponding auditory filters (Lin and Hartmann 1997). All stimuli were gated on and off with a 50-ms Hanning window. The stimuli were presented at an overall level of 67 dB sound pressure level.

### MEG Recordings and Data Analysis

The gradients of the AEF were recorded using a whole-head 122-channel gradiometer system (Neuromag Elekta Oy, Finland; Ahonen et al. 1993) inside a magnetically shielded room (Imedco, Switzerland) at a sampling rate of 1,000 Hz. During the recording, data were low-pass filtered at 330 Hz. Prior to data acquisition, the nasion, 2 preauricular points and 32 surface points were measured as individual anatomical landmarks for each participant using a Polhemus 3D Space Isotrack2 system. To maintain vigilance, subjects watched a silent movie of their own choice while the AEFs were collected. The subjects were instructed to concentrate on the movie and ignore the sounds in the earphones. During recording, 200 sweeps per stimulus condition were delivered to the subjects, diotically and in pseudorandomized order, via Etymotic Research (ER3) earphones with 90-cm plastic tubes and foam earpieces. The stimuli were played at a 48,000 Hz sampling rate using a 24-bit sound card (RME ADI 8DS AD/DA interface), an attenuator (Tucker-Davis Technologies PA-5) and a headphone buffer (Tucker-Davis Technologies HB-7).
The duration of the interstimulus interval varied randomly between 800 and 850 ms. The total duration of the MEG recording session was 32 min.

Data were analyzed offline using the BESA 5.2 software package (BESA, Germany) with a spherical head model and a homogeneous volume conductor. Artifacts were rejected by visual inspection of the raw data and by automatic rejection tools in BESA: first, noisy channels were excluded from further analyses; then, epochs with amplitudes $>8,000$ fT/cm or gradients $>800$ fT/cm were discarded. The average level in the interval of $-100$ to 0 ms, relative to stimulus onset, was used as the baseline.

In an effort to model the different components of the AEF in response to the overall onset, as well as the second stimulus segment, we applied spatiotemporal source models (Scherg 1990; Scherg et al. 1989; Scherg and Picton 1991) using BESA. In this source localization approach, equivalent current dipoles are modeled as intracerebral sources of the activity observed at the scalp, and the spatial position and orientation of the dipoles is iteratively varied until maximum variance is explained in the scalp data. The source model comprises the spatial information for each dipole and its physiological activity across time (source wave). Source models were calculated with one dipole per hemisphere, separately for each AEF component at the onset and at the transition, and for the steady-state response (SSR) during the first stimulus segment. Different band-pass filters were applied to carry out the dipole fits: for the N1m and P2m components, MEG data were zero-phase filtered 1–30 Hz; for the P1m component, a 2- to 60-Hz forward filter was used. The dipole fits on the SFs were based on unfiltered MEG data; for the SSR fit, data were zero-phase filtered from 20 to 100 Hz. For the onset AEF components, dipole fits were conducted on pooled conditions (SIN2 + SIN19 + RAND2 + RAND19); for the responses at the transition, fits were based on the SIN19 condition. The fitting interval was about 30 ms in duration around the peak of the P1m and N1m components. For the P2m component, the fitting interval covered about 50–70 ms around its peak. The SF was fitted in the interval from 500 to 800 ms after onset or transition, respectively, and the SSR was fitted from 300 to 800 ms after overall onset.

Individual fits for the onset AEF responses were successful for all 13 subjects in the case of the P1m and SF components, and for 8, 10 and 12 subjects in the case of the SSR, N1m, and P2m components, respectively. A symmetry constraint was introduced to stabilize the individual fits where necessary. For the P1m, N1m, P2m, and SSR dipoles at the onset, symmetry conditions were applied in two, two, five, and four subjects, respectively; for the onset SF dipoles, no symmetry constraint was needed for any subject. Turning to the AEF responses at the transition, dipole fits were successful for all 13 subjects in the case of the N1m and SF components, and for 12 subjects in the case of the P1m and P2m components, respectively. For the P1m and P2m dipoles, symmetry conditions were applied in four and three subjects, respectively; for the N1m and SF dipoles, no symmetry constraint was needed for any subject. The subjects who did not show stable fits in a dipole model were discarded from all further analyses regarding that model. No further constraints were introduced concerning the orientation and location of the dipoles. Table 1 shows means and standard deviations for the Talairach coordinates (Talairach and Tournoux 1988) of the dipole models from experiment 1, based on the standardized head model as implemented in BESA.

After fitting, the dipole models were used as spatiotemporal filters, and the source waves were derived for all four experimental conditions: i.e., the source waves corresponding to each dipole model were extracted separately for each condition. In the case of the SF models, a principal component analysis (Berg and Scherg 1994) was calculated over the last 50 ms of the epoch for each condition to compensate for drift. The component that explained the most variance was included in the model for that condition. Finally, all source waves were exported to MATLAB for statistical analysis. In the case of the P1m component, the source waves were low-pass filtered at 25 Hz prior to statistical analysis to avoid interference from the SSR.

Prior to statistical analyses, each source wave was adjusted to the baseline calculated as the average of the last 100 ms before stimulus onset or transition, respectively. MEG data were evaluated statistically using a bootstrap technique (Efron and Tibshirani 1993). In this robust and distribution-free method, the distribution of any test statistic can be approximated by repeated random drawing, with replacement, from the original data. The resulting bootstrap distribution is used to derive confidence intervals for this test statistic. In contrast to most standard methods, bootstrap techniques are well-suited for physiological datasets in which peaks cannot be clearly identified for each subject in every experimental condition, which was the case in our study. If, in the following, bootstrap comparisons are reported between a single condition and $n > 1$ other conditions, this does not mean $n$ pairwise comparisons; rather, it refers to one contrast between the single condition and the averaged data of those other conditions. A significance level of 1% was used throughout the paper to compensate for bias due to multiple testing.

### Auditory Models

The auditory models were hosted by AIM-MAT (Bleeck et al. 2004; http://code.soundsoftware.ac.uk/projects/aim), the MATLAB implementation of the auditory image model (e.g., Patterson et al. 1995). BMM was simulated separately for the two segments of each stimulus in both experiments; the sampling rate was 20,000 Hz. Precocchlear processing by the outer and middle ear was modeled with the loudness contour described by Glasberg and Moore (2002). Then, the BMM produced by each stimulus was simulated using both the GT-AFB described above, and a transmission line auditory filterbank (TL-AFB) proposed by Sieroka et al. (2006). Uppenkamp et al. (2001) implemented the auditory image model (e.g., Patterson et al. 2004; http://code.soundsoftware.ac.uk/projects/aim), the MATLAB implementation of the auditory image model (e.g., Patterson et al. 1995). BMM was simulated separately for the two segments of each stimulus in both experiments; the sampling rate was 20,000 Hz. Precocchlear processing by the outer and middle ear was modeled with the loudness contour described by Glasberg and Moore (2002). Then, the BMM produced by each stimulus was simulated using both the GT-AFB described above, and a transmission line auditory filterbank (TL-AFB) proposed by Sieroka et al. (2006). Uppenkamp et al. (2001)

### Table 1. Means and SD for the Talairach coordinates of the dipole models from experiment 1

<table>
<thead>
<tr>
<th></th>
<th>Left</th>
<th></th>
<th>Right</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$x$</td>
<td>$y$</td>
<td>$z$</td>
<td>$x$</td>
</tr>
<tr>
<td>Onset</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P1m</td>
<td>$-50.8 (4.6)$</td>
<td>$-27.0 (6.4)$</td>
<td>$6.6 (6.1)$</td>
<td>$48.3 (5.4)$</td>
</tr>
<tr>
<td>N1m</td>
<td>$-55.8 (3.3)$</td>
<td>$-31.1 (10.7)$</td>
<td>$4.3 (6.2)$</td>
<td>$53.1 (4.8)$</td>
</tr>
<tr>
<td>P2m</td>
<td>$-48.2 (5.3)$</td>
<td>$-20.9 (4.0)$</td>
<td>$-1.1 (6.8)$</td>
<td>$48.2 (6.3)$</td>
</tr>
<tr>
<td>SF</td>
<td>$-47.5 (3.1)$</td>
<td>$-22.7 (4.6)$</td>
<td>$0.7 (4.1)$</td>
<td>$45.4 (1.8)$</td>
</tr>
<tr>
<td>SSR</td>
<td>$-49.4 (6.2)$</td>
<td>$-24.2 (8.6)$</td>
<td>$10.4 (6.4)$</td>
<td>$47.3 (5.8)$</td>
</tr>
<tr>
<td>Transit</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P1m</td>
<td>$-51.5 (6.0)$</td>
<td>$-24.6 (6.1)$</td>
<td>$3.8 (9.8)$</td>
<td>$50.1 (6.6)$</td>
</tr>
<tr>
<td>N1m</td>
<td>$-47.8 (4.1)$</td>
<td>$-23.9 (6.3)$</td>
<td>$1.2 (4.5)$</td>
<td>$46.8 (3.7)$</td>
</tr>
<tr>
<td>P2m</td>
<td>$-46.7 (4.6)$</td>
<td>$-20.4 (4.2)$</td>
<td>$-0.7 (5.4)$</td>
<td>$45.2 (5.3)$</td>
</tr>
<tr>
<td>SF</td>
<td>$-46.5 (3.5)$</td>
<td>$-23.3 (3.5)$</td>
<td>$-1.0 (4.6)$</td>
<td>$43.6 (4.5)$</td>
</tr>
</tbody>
</table>

Values are means (SD). SF, sustained field; SSR, steady-state response.
RESULTS

Figure 2 depicts the MEG grand-mean source waves in response to the first stimulus segment (0–800 ms) in experiment 1. The two SIN conditions and the two RAND conditions have been pooled, separately, because the first stimulus segment was identical in each case. The data have also been pooled across hemispheres, since no overall lateralization effects were found for either AEF component at the onset or the transition (the P values obtained from the statistical evaluation of the AEF components are given in Table 2). The overall onset of the tone complex elicited AEF responses with P1m, N1m, and P2m components as well as subsequent SFs in all conditions. No significant amplitude differences occurred between SIN and RAND conditions for any of the onset AEF components; however, the RAND stimuli elicited P1m and N1m responses with slightly shorter latencies than the SIN stimuli. The magnitude of the SF at the end of the first interval was the same for these average SIN and RAND source waves.

In addition to the standard AEF components, there was a distinct SSR at 50 Hz (f0) riding on the SF in the source waves. It was observed in all experimental conditions, and it was present throughout the stimulus. Moreover, the magnitude of the SSR appeared to be stronger in the SIN conditions than in the RAND conditions. Accordingly, a fast Fourier transform was performed on the source waves of the SSR model (0–800 ms, zero-padded to 1,000 ms), and the magnitudes of the 50-Hz peaks in the resulting power spectra were evaluated with the bootstrapping technique. This procedure revealed that the SSR was more pronounced in the SIN conditions than in the RAND conditions (cf. Table 2).

The neuromagnetic response to the emergence of DP in the second half of the stimulus interval is depicted in Fig. 3, which shows the grand-average source waves of experiment 1 for the complete stimulus duration; the waves were high-pass filtered at 25 Hz to remove the SSR. There is a clear onset complex (P1m, N1m and P2m) at the transition to the second stimulus segment (at 800 ms) for the SIN19 condition but not for any of the other conditions (cf. Table 2). It is also the case that the SF in the SIN19 condition becomes more negative when the 19th harmonic (950 Hz) is removed compared with the SF in the other conditions (cf. Table 2). This is also the case that the SF in the SIN19 condition becomes more negative when the 19th harmonic (950 Hz) is removed compared with the SF in the onset segment of the same condition (P < 0.001), and there is no corresponding effect in the source waves of the other conditions. These transient and sustained deviations of the MEG in the SIN19 condition are neuromagnetic evidence of DP.

Table 2. P values obtained from the statistical evaluation of the auditory evoked field components in experiment 1

<table>
<thead>
<tr>
<th></th>
<th>P1m Lat</th>
<th>P1m Amp</th>
<th>N1m Lat</th>
<th>N1m Amp</th>
<th>P2m Lat</th>
<th>P2m Amp</th>
<th>SF Amp</th>
<th>SSR 50-Hz Peak</th>
</tr>
</thead>
<tbody>
<tr>
<td>Onset</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left vs. right</td>
<td>0.05</td>
<td>0.27</td>
<td>0.36</td>
<td>0.41</td>
<td>0.09</td>
<td>0.32</td>
<td>0.37</td>
<td>0.37</td>
</tr>
<tr>
<td>SIN2 + SIN19 vs. RAND2 + RAND19</td>
<td>&lt;0.001*</td>
<td>0.40</td>
<td>&lt;0.001*</td>
<td>0.19</td>
<td>0.04</td>
<td>0.08</td>
<td>0.44</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>Transition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left vs. right</td>
<td>0.27</td>
<td>0.18</td>
<td>0.40</td>
<td>0.19</td>
<td>0.49</td>
<td>0.10</td>
<td>0.47</td>
<td></td>
</tr>
<tr>
<td>SIN19 vs. (SIN2 + RAND2 + RAND19)</td>
<td>0.01</td>
<td>&lt;0.001*</td>
<td>0.13</td>
<td>&lt;0.001*</td>
<td>&lt;0.001*</td>
<td>&lt;0.001*</td>
<td>&lt;0.001*</td>
<td></td>
</tr>
<tr>
<td>SIN2 vs. (RAND2 + RAND19)</td>
<td>0.10</td>
<td>0.23</td>
<td>0.04</td>
<td>0.16</td>
<td>0.32</td>
<td>0.20</td>
<td>0.12</td>
<td></td>
</tr>
</tbody>
</table>

Amp, amplitude; Lat, latency; SIN, sine phase; RAND, random phase. For each component, the baseline of the source waves was set to the 100-ms interval before the onset or the transition, respectively. *Significant difference: P < 0.001.

J Neurophysiol • doi:10.1152/jn.00898.2013 • www.jn.org
The BMM plots in the first two rows of Fig. 1 provide a reasonable explanation of the DP findings. There are quiet intervals in the higher frequency channels of the SIN19 BMM, and the oscillation that arises when the 19th harmonic is removed is readily observed in the quiet period (right panel of top row). There are no corresponding oscillations in the BMM of the SIN2 stimulus because there are no quiet intervals in low-frequency channels. For the RAND BMMS in row 2, there are no quiet intervals around the pitch ridge, and so it is not possible to observe any oscillations associated with absent harmonics in either the RAND2 or RAND19 conditions. The results support the peripheral-origin hypothesis that a DP is heard when the removal of a high harmonic produces an oscillation on the basilar membrane in a quiet interval within the period of a MP sound.

In the psychoacoustic experiment, the perceived change at the transition to the second stimulus segment was found to be much more detectable in the SIN19 condition than in the other conditions [(SIN2 + RAND2 + RAND19); P < 0.001]. This was expected because it is this condition that produces a clear DP percept when the 19th harmonic (950 Hz) is omitted. In the SIN2, RAND2 and RAND19 sounds, it was difficult to perceive any change at the onset to the second segment of the HCT.

Interim Summary

The MEG data from experiment 1 clearly reflect peripheral auditory processing as well as the perception of the stimuli. In the SIN19 condition, at the transition to the second stimulus segment, the emergence of DP in the perception is accompanied by a pronounced onset complex (P1m, N1m and P2m) and a subsequent SF increment (see Fig. 3). These AEF responses can be regarded as DP specific, since they are not elicited by the omission of a low harmonic or the associated energy reduction; rather, they only occur when the subjects hear a DP join the ongoing MP.

At the onset of the stimulus, the SIN and RAND stimuli all produce onset complexes (P1m, N1m and P2m) with the standard form. The RAND stimuli have a distinctly ragged timbre (Patterson 1987) and a slightly weaker MP (Lundeen and Small 1984; Shofner and Selas 2002), but there is no obvious correspondence between the morphology of the onset complex and either the pitch or timbre of the stimuli. It is worth noting, however, that these deflections include the energy onset response typical of neuromagnetic responses to sound (Biermann and Heil 2000; Krumholz et al. 2003).

EXPERIMENT 2

The second experiment was designed to investigate the influence of stimulus phase spectrum on DP and the AEF in more detail. Specifically, two stimuli with phase spectra that produce pulsive waves (SIN and COS) were compared with two stimuli with phase spectra that produce minimum peak factor waves, namely, Schroeder-phase negative (SCHR −) and Schroeder-phase positive (SCHR +) stimuli. Despite the low peak factor, the BMM responses associated with SCHR − and SCHR + waves both exhibit quiet intervals at some point within the period in the higher frequency channels. The form of the quiet interval varies with SCHR sign, as shown in the bottom two rows of Fig. 1. SCHR sign is also known to affect the AEF in auditory cortex (Rupp et al. 2008).

MATERIALS AND METHODS

Participants

Eighteen normal-hearing adult listeners (12 women; all right-handed; mean age: 24.2 ± 5.8 yr) volunteered for experiment 2 after providing written, informed consent. None of the participants reported any history of peripheral or central hearing impairment or any psychiatric or neurological disorder.

Stimuli

All four stimuli (SIN, COS, SCHR −, and SCHR +) had the same level, f0, duration, gating functions and number of harmonics as in experiment 1. The same pop-out paradigm was employed as in experiment 1: in the first segment, all harmonic components were present; in the second segment, the 19th component (950 Hz) was removed to evoke a DP. There were also two control stimuli in the experiment; they did not reveal any interesting differences between conditions, so they were excluded from the analysis for brevity and clarity.

MEG Recordings and Data Analysis

All procedures regarding the acquisition and preprocessing of MEG data were the same as those described in experiment 1. The duration of the MEG recording session in experiment 2 was 48 min.

Separate spatiotemporal source models, each with one equivalent current dipole per hemisphere, were employed to analyze the specific neuromagnetic responses to overall onset and the transition to the second stimulus segment. All dipole fits were based on MEG data averaged across all four conditions. With respect to the onset responses, individual dipole modeling was successful for all 18 subjects in the case of the N1m and SF components, for 17 subjects in the case of the P1m and P2m components, and for 12 subjects in case of the onset SSR. To stabilize the fits, a symmetry constraint was applied in 3, 10, 6, and 5 subjects, respectively, for the P1m, N1m, P2m, and SF dipoles. For the SF dipoles, no symmetry constraint was needed in any subject. Similarly, fitting the transition responses was successful for all 18 subjects in the case of the SF component, and for 17 subjects in the case of the P1m, P1m, N1m, and P2m components. For the P1m, N1m, and P2m dipoles at the transition, symmetry was applied in three, nine, and two subjects, respectively, for the SF dipoles, no symmetry constraint was required. The subjects who did not show stable fits in a dipole model were excluded from all analyses regarding this model. With respect to the location and orientation of the dipoles, no further constraints were introduced. All procedures regarding the derivation and drift-correction of the source waves and their statistical analyses were identical to experiment 1. Table 3 shows means and standard deviations for the Talairach coordinates of the dipole models.
Table 3. Means and SD for the Talairach coordinates of the dipole models from experiment 2

<table>
<thead>
<tr>
<th></th>
<th>Left</th>
<th>Right</th>
</tr>
</thead>
<tbody>
<tr>
<td>Onset</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P1m</td>
<td>50.4 (7.9)</td>
<td>52.3 (5.6)</td>
</tr>
<tr>
<td>N1m</td>
<td>51.8 (7.8)</td>
<td>53.7 (6.8)</td>
</tr>
<tr>
<td>P2m</td>
<td>-48.4 (7.0)</td>
<td>46.7 (7.5)</td>
</tr>
<tr>
<td>SF</td>
<td>-45.8 (5.9)</td>
<td>45.9 (5.4)</td>
</tr>
<tr>
<td>SSR</td>
<td>-52.6 (11.3)</td>
<td>52.2 (8.6)</td>
</tr>
<tr>
<td>Transition</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P1m</td>
<td>-51.8 (7.7)</td>
<td>48.8 (5.5)</td>
</tr>
<tr>
<td>N1m</td>
<td>-51.7 (4.9)</td>
<td>50.4 (6.6)</td>
</tr>
<tr>
<td>P2m</td>
<td>-49.3 (6.2)</td>
<td>47.8 (6.9)</td>
</tr>
<tr>
<td>SF</td>
<td>-44.0 (6.5)</td>
<td>44.1 (6.2)</td>
</tr>
</tbody>
</table>

Values are means (SD).

from experiment 2, based on the standardized head model as implemented in BESA.

RESULTS

Figure 4A shows the MEG source waves in response to the first segment of each stimulus in experiment 2; the statistics regarding differences in components of the AEF are presented in Table 4. The data were pooled across hemispheres, since no laterization effects occurred for any AEF component at the onset or at the transition to the second segment, with the exception of the SSR, which was somewhat greater in the right hemisphere.

As in experiment 1, the source waves exhibited pronounced SSRs, corresponding to f0, throughout the duration of the stimulus. This oscillation obscures the form of the onset response and the subsequent DP response. Accordingly, we describe it first and then remove it. A fast Fourier transform was used to measure the magnitude of the SSR; the spectral power in the region from 47- to 52-Hz region is shown in Fig. 4B. The analysis shows that the SIN and COS stimuli produced very similar SSR magnitudes, while the SCHR− condition produced a larger SSR than the remaining conditions [(SIN + COS + SCHR+); P < 0.01] and the SCHR+ condition produced a smaller SSR than the remaining conditions [(SIN + COS + SCHR−); P < 0.001]. The thin vertical lines at 600 ms in Fig. 4A show the relative phases of the SSR components of the source waves. They are the same for the SIN and COS stimuli, and they differ from the relative phases of both the SCHR− and SCHR+ stimuli, indicating that fine details of the 50-Hz oscillation are preserved in auditory cortex.

Figure 5 presents the grand-mean source waves for the four stimuli of experiment 2 for the complete stimulus duration, after the waves were high-pass filtered at 25 Hz to remove the SSR. The figure shows that all of the stimuli elicited tripartite onset responses and a large SF. In the first segment (up to 800 ms), the SIN and COS stimuli with large peak factors produce source waves that are very similar in form, and the SCHR stimuli with minimum peak factors produce source waves that are very similar in form, despite the pronounced differences in the BMMs they produce. The SCHR stimuli elicit P1m’s with slightly shorter latencies than the SIN and COS stimuli, and they produce SFs with greater magnitudes than the SIN and COS stimuli, but the SCHR source waves do not differ as a function of the chirp direction of the stimulus (cf. Table 4). It seems that the factors which are thought to explain SSR magnitude (i.e., modulation depth within channels and temporal alignment across channels) are not the determining factors for the more central AEF components (in this case, the P1m and the SF). It may be that the SCHR waves with their frequency glides place a greater processing burden on the auditory system, or recruit more neurons during the extraction of auditory features associated with pitch and timbre. In any event, the slow AEF components are not sensitive to peripheral phase effects in the way that the SSR is.

In this experiment, in the second segment of the stimulus, it is invariably the 19th harmonic that is removed, and all of the stimuli have orderly phase spectra in the region of the 19th harmonic. As a result, they all produce a DP perception, and...
the source waves all exhibit a tripartite “DP onset” response following the removal of the 19th harmonic. In the SIN and COS source waves, these DP onset responses are very similar, and both are similar to the onset response for the SIN19 source wave in experiment 1, although the N1m magnitude is somewhat reduced in experiment 2. It is also the case that the DP onset response in the SCHR+ source wave is similar in form to that for the SIN and COS waves; there are no significant differences in peak delay or amplitude for any of the onset components (cf. Table 4). However, the DP onset response of the SCHR− stimulus is different; the amplitude of the P1m in the SCHR− source wave is significantly smaller than those in the SIN, COS and SCHR+ waves, and the peak delays of the N1m and P2m are significantly longer in the SCHR− source wave than in the SIN, COS and SCHR+ source waves.

Since the SCHR waves produce very similar onset responses at the start of the first segment, it seems likely that the source-wave differences between the SCHR− and SCHR+ responses at the start of the second segment reflect differences in DP processing or DP perception. It was noted in the Introduction that the BMM plots in Fig. 1 suggest that DP will be less audible in the SCHR− condition than in the SCHR+ condition (and the SIN condition). Accordingly, a discrimination experiment was performed in five normal-hearing volunteers (mean age: 31.2 yr) to determine the relative strength of the perceptual cue associated with DP in three of the conditions of experiment 2 (SIN, SCHR+ and SCHR− phase). The stimuli and earphones were the same as those used for the paired comparison (BTL) experiment in Psychoacoustics above.

On each trial, the listener was presented with two versions of the stimulus which had all harmonics present and one version in which the 19th harmonic (950 Hz) was attenuated by x dB; the order of the three stimuli was randomized on each trial, and the listener was asked to indicate which of the three sounds contained a DP. Between trials, x was varied adaptively in a 3-down, 1-up procedure (Levitt 1971) to determine the attenuation required to support reliable detection of the interval with the attenuated 19th harmonic. The initial attenuation was 8 dB to make the task easy at the start of the run, and the initial step size was 2 dB. After two reversals the step size was reduced to 1 dB, and after another four reversals it was reduced to 0.5 dB. The run was terminated after 13 reversals, and DP threshold was taken to be the average value across the last six reversals.

Each listener performed two runs for each phase configuration; the first run served as practice, while the second run was used to determine threshold. Feedback was provided at the end of each trial.

The results were straightforward: DP becomes audible in the SIN and SCHR+ conditions when the attenuation is only about 2.5 dB, whereas the attenuation has to be greater than 6 dB, on average, to support DP perception in the SCHR− condition. The threshold difference between the SCHR− and SCHR+ conditions is highly significant ($P < 0.001$). The listeners also commented that it was difficult to hear the DP in the SCHR− condition. In summary, the DP onset responses that occur when the 19th harmonic is removed from SCHR waves are compatible with the peripheral origin hypothesis for DP. The SCHR− wave, which reduces the depth of modulation in the BMM, and which makes the DP feature harder to detect, leads to a source-wave onset complex with reduced amplitude and increased delay.

In the latter part of the second segment, after the P2m, the source waves of the SCHR− and SCHR+ stimuli converge on the same SF value, and that value is the same as these stimuli produced in the latter part of the first interval where the 19th harmonic was present (not shown). As the interval proceeds beyond 1,300 ms, the SF appears to decay slightly, but, for both of the waves with low peak factors, the SF remains considerably larger (more negative) than that produced by the waves with large peak factors. This suggests that the magni-

---

### Table 4. $P$ values obtained from the statistical evaluation of the auditory evoked field components in experiment 2

<table>
<thead>
<tr>
<th>Component</th>
<th>Lat</th>
<th>Amp</th>
<th>Lat</th>
<th>Amp</th>
<th>Lat</th>
<th>Amp</th>
<th>SF Amp</th>
<th>SSR 50-Hz Peak</th>
</tr>
</thead>
<tbody>
<tr>
<td>Onset</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left vs. right</td>
<td>0.44</td>
<td>0.29</td>
<td>0.65</td>
<td>0.45</td>
<td>0.14</td>
<td>0.42</td>
<td>0.22</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>SCHR− vs. SCHR+</td>
<td>0.02</td>
<td>0.53</td>
<td>0.22</td>
<td>0.49</td>
<td>0.01</td>
<td>0.36</td>
<td>0.49</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>(SCHR− + SCHR+) vs. (SIN + COS)</td>
<td>&lt;0.001*</td>
<td>0.25</td>
<td>0.03</td>
<td>0.15</td>
<td>0.08</td>
<td>0.37</td>
<td>&lt;0.001*</td>
<td>0.25</td>
</tr>
<tr>
<td>Transition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left vs. right</td>
<td>0.31</td>
<td>0.10</td>
<td>0.44</td>
<td>0.33</td>
<td>0.52</td>
<td>0.35</td>
<td>0.33</td>
<td></td>
</tr>
<tr>
<td>(SCHR− + SCHR+) vs. (SIN + COS)</td>
<td>0.12</td>
<td>&lt;0.01†</td>
<td>&lt;0.01†</td>
<td>0.18</td>
<td>0.09</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCHR− vs. (SIN + COS)</td>
<td>0.04</td>
<td>&lt;0.01†</td>
<td>&lt;0.001*</td>
<td>0.06</td>
<td>0.34</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCHR+ vs. (SIN + COS)</td>
<td>0.05</td>
<td>0.09</td>
<td>0.01</td>
<td>0.23</td>
<td>0.02</td>
<td>0.48</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>SCHR− vs. SCHR+</td>
<td>0.05</td>
<td>&lt;0.001*</td>
<td>0.04</td>
<td>0.42</td>
<td>0.43</td>
<td>0.06</td>
<td>0.07</td>
<td></td>
</tr>
</tbody>
</table>

SCHR−, Schroeder phase instantaneous frequency increase; SCHR+, Schroeder phase instantaneous frequency decrease; COS, cosine phase. For each component, the baseline of the source waves was set to the 100-ms interval before the onset or the transition, respectively. Significant differences: *$P < 0.001$, †$P < 0.01$.

---

![MEG grand-mean source waves of experiment 2](http://jn.physiology.org/)

**Fig. 5.** MEG grand-mean source waves of experiment 2. The data shown here are based on a simultaneous dipole fit at the onset to the first and second segments. The data were pooled across hemispheres and low-pass filtered at 25 Hz to provide a better overview of the amplitude dynamics for the complete stimulus duration.
tude of the SF is more closely associated with the processing of the HCT than the extraction or perception of the DP.

Finally, in the second segment, the SF appears to vary in magnitude between experimental conditions (Fig. 5, 1,200–1,600 ms). In an effort to reveal any significant differences, we created a new dipole model for the sustained response extending throughout the entire stimulus. Modeling was conducted in BESA in the same manner as described in MEG Recordings and Data Analysis above; but here, the dipole fit was made for the SF at onset and transition simultaneously rather than separately. Individual fits were successful in all participants; a symmetry constraint was not needed in any case. No significant SF increment was found for the pooled SCHR conditions \([P = 0.26, \text{nonsignificant (ns)}]\), nor for the pooled SIN and COS conditions \([P = 0.05, \text{ns})\]. To test for interactions, we subtracted the onset and transition SFs within the respective groups of conditions and evaluated the corresponding difference waves with bootstrapping. Here, the interaction between the SCHR conditions and the remaining conditions was not significant \([\text{SCHR}− + \text{SCHR}+] \text{vs. (SIN + COS)}; P = 0.02, \text{ns}]\.

DISCUSSION

The experiments show that there is a pronounced neuromagnetic response in auditory cortex when listeners hear a DP and the form of the source wave is affected by the phase spectrum of the HCT that produces the initial complex tone with its MP.

Source Wave Morphology in the First Segment of the Stimulus

The grand-mean source waves observed in response to the HCT in the first segment of each trial provide the initial data from these experiments on the relationship between auditory perception and components of the AEF in auditory cortex. In both experiments, the responses exhibit standard onset complexes consisting of a P1m, an N1m and a P2m which then evolves into a large SF about 400 ms after stimulus onset. The differences between the responses, which reflect differences in the phase spectra of the stimuli, appear in \(I\) the magnitude and phase of the SSR riding on the source wave (Figs. 2 and 4), \(J\) the magnitude of the SF itself, and \(J\) the details of the onset response (Figs. 3 and 5).

SSR. With regard to the SSR, in both experiments, the magnitude of the 50-Hz oscillation is related to the peak factor of the stimulus wave: in experiment 1 (Fig. 2), the SSR magnitude is considerably larger for SIN waves which have a large peak factor, than for RAND waves which have a small peak factor. In experiment 2 (Fig. 4), the SIN and COS waves with large peak factors also produce pronounced SSRs whose magnitudes are similar to those for the SIN waves in experiment 1. The low peak factor SCHR+ wave produces an SSR whose magnitude is small relative to those of the SIN and COS waves. But the SCHR− wave produces an SSR whose magnitude is, if anything, larger than those of the large peak factor waves (SIN and COS).

The obvious explanation in this case is that the SCHR− wave tends to counteract the natural phase delay in the cochlea, which increases the number of filter channels whose modulation cycles go through their peaks and troughs at the same time (Kohlrausch and Sander 1995; Rupp et al. 2008). The results suggest that SSR magnitude is closely related to stimulus peak factor and phase alignment in the cochlea, that is, to peripheral aspects of auditory processing. This interpretation is reinforced by the relative phases of the SSSRs. The small vertical lines at 600 ms in Fig. 5 show that there are consistent differences in phase between the SSRs. It is surprising to see such a precise representation of cochlear phase in a cortical response, especially as the source waves are grand averages over replications, both within a listener and over listeners as well. It is this precise encoding of phase that makes it possible to observe the enhancement of phase alignment produced by the SCHR− wave in the cochlea and the reduction in phase alignment by the SCHR+ wave in auditory cortex.

Onset response. The source waves of Figs. 2 and 4 are replotted in Figs. 3 and 5 after low-pass filtering to remove the SSR and provide a better representation of the tripartite onset responses (and the SFs) produced by these complex MP tones (the region up to 800 ms). Comparison of the source waves both within and across experiments shows that the onset response and the SF are not dominated by modulation depth and alignment in the cochlea as in the case of the SSR. In experiment 1, the delays of the P1m and N1m peaks are slightly shorter in the RAND conditions than they are in the SIN conditions, but this is the only significant difference between the SIN and RAND source waves; the major differences in the BMM patterns produced by SIN and RAND stimuli (Fig. 1) are not reflected in the source waves at stimulus onset (cf., Table 4).

In experiment 2, the SCHR+ and SCHR− waves produce very similar P1m-N1m complexes, despite their differential effects in the cochlea; moreover, the amplitude difference in the latter part of the P2m is not significant. The delay of the P1m in the pooled SCHR+ and SCHR− response is a little shorter than in the pooled SIN and COS response, but this just appears to be part of the progressive divergence of the pooled responses which eventually culminates in a larger SF for the low peak-factor waves. The major differences in the BMM patterns produced by SCHR+ and SCHR− waves (Fig. 1) are not reflected in the source waves at stimulus onset.

SF. In experiment 1, as the onset complex ends and the SF forms, the RAND source waves become indistinguishable from the SIN source waves (cf., Table 2). Despite the major differences in BMM patterns, the SF of the source wave is the same for the two classes of stimuli throughout the first segment. The RAND stimuli of experiment 1 are readily discriminable from the COS and SIN stimuli on the basis of their ragged timbre (Patterson 1987). They also have a somewhat weaker pitch (Shofner and Selas 2002). Nevertheless, the SF is the same for all of these stimuli. (Similarly, the minor differences in the onset response do not bear any obvious relationship to the perceptual differences.)

In experiment 2, the SIN and COS waves with large peak factors produce SFs with the same magnitude, as would be expected, and the magnitude is about the same size as for the SIN waves in experiment 1. However, the SCHR waves with minimal peak factors both produce larger (more negative) SFs, and they produce the same SF, despite their differential effects on modulation depth and alignment in the cochlea. Thus the SF is not like the SSR; its magnitude is not dominated by modulation depth and alignment in the cochlea.

\[ JNeurophysiol \text{ \cdot doi:10.1152/jn.00898.2013 \cdot www.jn.org} \]
The fine structure of the SCHR waves chirps across the 20-ms period of the wave. But the acoustic chirp does not impart a corresponding up chirp to the SCHR− sound or a down chirp to the SCHR+ sound; the period of the wave is too short to support the perception of a repeating frequency glide. Both stimuli are perceived to have a slight flutter, or instability, which contrasts with the static perception that accompanies the SIN and COS stimuli. The difference in perceived stability might be associated with the difference in SF magnitude (or P1m latency); perhaps it takes more effort to process the repeating frequency glides of the SCHR stimuli. But if this were the case, we might have expected the RAND stimuli in experiment 1 to produce larger SFs than the SIN stimuli, which they do not. Thus there does not seem to be any simple perceptual interpretation of the differences between the source waves of the high and low peak factor waves. This suggests that the morphology of the onset complex and SF reflect processing activity in more central stages of the auditory pathway. This is not particularly surprising; physiological data from awake primates show that frequency modulated sinusoids produce greater sustained responses than unmodulated sinusoids in primary auditory cortex (Liang et al. 2002, in vivo extracellular recordings), while there is no corresponding difference in auditory nerve fibers (Sinex and Geisler 1981). Nevertheless, it is not immediately clear what features of the stimuli produce the distinguishing features of the source waves.

**Source Wave Morphology in the Second Segment**

At the end of the first interval of each trial, the HCT has invariably established a large stable SF that provides a platform for observing the onset of DP. In the first experiment, the removal of one harmonic produces a deflection of the SF in one and only one case, when the 19th harmonic (950 Hz) is removed from the SIN wave. This is also the one stimulus in this experiment that produces a DP. The response consists of an onset complex (P1m, N1m, P2m) whose magnitude is comparable to that produced by the HCT at the onset of the first segment. There is also a small increase in the magnitude of the SF relative to the SFs produced by the other three stimuli. The difference is largest at 1,300 ms and decays back toward the average SF as the interval proceeds to 1,600 ms (not shown). Together, the onset complex and the enhanced SF represent the MEG response to the DP that listeners hear. The response appears when the filtering action on the basilar membrane causes the membrane to vibrate in the region of the 19th harmonic when it would otherwise be quiescent. Thus the source waves from experiment 1 support the peripheral origin hypothesis.

In experiment 2, when the 19th harmonic is removed at the end of the first segment, all of the stimuli produce a DP perception, and the source waves all exhibit a clear DP onset response. The response to the SCHR+ wave is similar in form to that for the SIN and COS waves; the BMM pattern for the SCHR+ wave contains a broad quiet section in all of the high-frequency channels. The response to the SCHR− wave is different: the amplitude of the P1m is significantly smaller than those for the SIN, COS and SCHR+ responses, and the peak delays of the N1m and P2m are significantly longer in the SCHR− response than in the responses to the SIN, COS and SCHR+ waves. The DP is also less detectable in the SCHR− condition. Thus the DP onset responses are compatible with the hypothesis that DP is of peripheral origin, in the sense that they reflect the relative magnitude of the oscillation associated with the missing harmonic in the BMM patterns. The SCHR− wave, whose BMM has less modulation depth, and which produces a less detectable DP, leads to a source wave onset complex with reduced amplitude and increased delay. This finding is compatible with previous MEG work on masking period patterns obtained with SCHR stimuli (Rupp et al. 2008); the sinusoidal signal, which was more detectable in the SCHR+ masker than in the SCHR− masker, was observed to produce a stronger MEG response (P1m) in SCHR+ conditions.

**Peripheral Models of DP**

A comparison is presented in Fig. 6 of the simulated patterns of BMM produced by the TL-AFB and the GT-AFB in response to the stimuli of experiments 1 and 2. Both models provide support for the peripheral origin hypothesis: there is a DP-specific feature in the form of a high-frequency oscillation in the quiet period of the BMM in both panels of row 1 (the SIN stimulus), a condition where DP is clearly audible. There are no corresponding features in the panels of row 2 (the RAND stimulus), a condition where DP is not audible. Similarly, there is a DP-specific oscillation in the quiet period of the BMM in both panels of row 4 (the SCHR+ stimulus), the SCHR condition where DP is clearly audible. Finally, there is a prominent DP-specific oscillation in the BMM of the SCHR+ stimulus when it is generated with the GT-AFB (row 3, right column). The oscillation is as prominent as it is in the panel below for the SCHR+ stimulus with the GT-AFB. Perceptually, however, DP was found to be less audible with the SCHR− stimulus than with the SCHR+ stimulus. The perceptual difference is more compatible with the BMMs produced by the TL-AFB, - the panels in the left column of rows 3 and 4. The DP-specific oscillation is identifiable in the BMM of the SCHR− stimulus, but it is much less prominent than in the BMM of the SCHR+ stimulus. Thus peripheral models can explain the perception of DP and the main features of MEG source waves, but the more physiologically accurate, transmission-line model is required to explain the difference in DP strength between the SCHR− and SCHR+ stimuli, and the differences between the MEG source waves that they produce.

Towards the end of their paper, Lin and Hartmann (1997) describe an intriguing experiment concerning the phase spectrum of the HCT used to generate DP: in the frequency region around the harmonic to be removed, the harmonics of the HCT were fixed in phase, as usual, but, outside this band, the phases of the harmonics were randomized. They varied the width of the band of coherent harmonics and found that it had to be surprisingly wide, 12 or more harmonics for listeners to hear the DP. They point out that this is not consistent with the BMM patterns produced by a GT-AFB, where the DP oscillation is restricted to the channel centered on the frequency of the harmonic that was removed. The BMM pattern produced by the GT-AFB for the SIN stimulus (right column, row 1) is very similar to that presented by Lin and Hartmann (1997), and it shows the DP-specific oscillation restricted to one or two channels near the 19th harmonic. Consequently, they con-
cluded that the GT-AFB fails as a peripheral model of DP, at least in this one respect.

The TL-AFB includes the coupling of basilar membrane sections observed in the cochlea; this is important for accurate representation of BMM when sounds have complex phase spectra (Uppenkamp et al. 2001). Comparison of the DP oscillations in the left and right columns of Fig. 6 shows that the coupling in the TL-AFB results in the DP feature appearing in the channel centered on the 19th harmonic and a range of channels extending some distance above the 19th harmonic. This could explain why there has to be a relatively wide region of phase coherence to produce a DP. In the band narrowing experiment of Lin and Hartmann (1997), as the region of coherence is reduced, it will interfere with the DP feature in the BMM of the TL-AFB sooner than in the BMM of the GT-AFB. Thus a peripheral model based on a TL-AFB might well be able to explain DP in its entirety. Since the broadening of the DP feature is limited to channels above the 19th harmonic, the TL-AFB predicts that the coherence band will be found to be asymmetric, extending further above the channel of the removed harmonic than below it. The coherence bands in Lin and Hartmann were not, however, sufficiently asymmetric to test this hypothesis.

The peripheral models also explain why the magnitude of the SSR is greater for the SCHR− stimulus than for the SCHR+ stimulus. For both stimuli, the frequency glide extends across the entire 20-ms period of the wave, either going from low to high (SCHR−) or from high to low (SCHR+). These glides interact with the normal phase delay in the cochlea which goes from high to low. As a result, the SCHR− wave tends to counteract the cochlear delay and concentrate neural firing within the period (row 3 of Fig. 6), whereas the SCHR+ wave tends to accentuate the cochlear delay and distribute neural firing more evenly across the period (row 4 of Fig. 6). The magnitude of the SSR reflects the distribution of firing across the period of the stimulus.

**Conclusion**

This paper describes cortical representations of pitch and phase information, and their relationships with peripheral auditory processing and perception. The MEG experiments show, for the first time, specific neuromagnetic activity associated with the onset of DP; moreover, they demonstrate how cortical responses to DP are affected by the phase spectrum of the stimulus, cochlear phase alignment, and higher auditory processes. Computer simulations of the cochlear mechanics can account for major portions of the neurophysiological and psychoacoustic results, both with respect to DP and the underlying HCT, provided that they include physiologically accurate coupling between adjacent frequency channels.

Fig. 6. Simulated BMM produced by the transmission line auditory filterbank (TL-AFB; left) and gammatone auditory filterbank (GT-AFB; right) in response to the second segment of the stimuli of experiments 1 and 2. There is a DP-specific high-frequency oscillation in the quiet period of each panel, with the exception of those associated with the RAND HCT. The coupling in the TL-AFB results in the DP feature appearing in a range of channels at and above the 19th harmonic.
ACKNOWLEDGMENTS

The authors thank H. G. Dosch for helpful discussions concerning basilar membrane models.

GRANTS

A scholarship provided by the Departments of Neurology and Otorhinolaryngology, University Hospital of Heidelberg (initiated by Prof. W. Hacke), is gratefully acknowledged by the first author.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author(s).

AUTHOR CONTRIBUTIONS


REFERENCES


