



ELSEVIER

Hearing Research 100 (1996) 164–170

HEARING
RESEARCH

Binaural fusion and the representation of virtual pitch in the human auditory cortex

Christo Pantev^a *, Thomas Elbert^b, Bernhard Ross^a, Carsten Eulitz^a, Ernst Terhardt^c

^a Center of Biomagnetism, Institute of Experimental Audiology, University of Münster, Kardinal-von-Galen-Ring 10, D-48129 Münster, Germany

^b University of Konstanz, D-78434 Konstanz, Germany

^c Technical University of Munich, D-80290 Munich, Germany

Received 6 April 1996; revised 17 June 1996; accepted 23 June 1996

Abstract

The auditory system derives the pitch of complex tones from the tone's harmonics. Research in psychoacoustics predicted that binaural fusion was an important feature of pitch processing. Based on neuromagnetic human data, the first neurophysiological confirmation of binaural fusion in hearing is presented. The centre of activation within the cortical tonotopic map corresponds to the location of the perceived pitch and not to the locations that are activated when the single frequency constituents are presented. This is also true when the different harmonics of a complex tone are presented dichotically. We conclude that the pitch processor includes binaural fusion to determine the particular pitch location which is activated in the auditory cortex.

Keywords: Hearing; Pitch perception; Binaural fusion; Magnetoencephalography; Auditory evoked magnetic field; Human auditory cortex

1. Introduction

The vowels of human speech, and the sounds produced by many musical instruments, are harmonic complex tones, i.e., comprised of pure tones the frequencies of which are integer multiples of a fundamental frequency. While the pitch of a pure tone is a primary auditory percept which is monotonically related to its frequency (spectral pitch), the pitch of complex tones ordinarily is of the virtual-pitch type. Both types of pitch can be assessed by aural matching to a pure tone. The virtual pitch of a harmonic complex tone largely corresponds to its fundamental frequency (oscillation frequency), even though the fundamental (the first harmonic) may be missing in the Fourier spectrum. According to the 'virtual-pitch theory' (Terhardt, 1972) virtual pitch is extracted by a type of 'Gestalt' recognition. Virtual pitch is considered as an attribute that is

established on a higher level of processing than spectral pitch. The distinction between spectral pitch and virtual pitch can be likened to the distinction between primary visual contour and virtual ('illusory') visual contour (Terhardt, 1987, 1989). Just as visual virtual contour can be regarded as being 'inferred' from a set of primary contours, virtual pitch is regarded as the result of a 2-stage process which involves peripheral frequency analysis and virtual-pitch formation. The formation of virtual pitch, then, requires the existence of a pertinent 'processor', and the position of that processor in the auditory pathway must be expected to be more 'central' than that pertinent to spectral pitch. This view is strongly supported by a number of psychoacoustic experiments the results of which demonstrate that the two ears intimately and readily cooperate in the formation of virtual pitch (Cramer and Huggins, 1958; Durlach, 1962; Houtsma and Goldstein, 1972; Goldstein, 1973; Bilsen and Goldstein, 1974; Van den Brink, 1974; Bilsen, 1976, 1977; Houtsma, 1979; Klein and Hartmann, 1981). The results of biomagnetic experiments described in the present paper provide

* Corresponding author. Fax: +49 (251) 83-6882;
E-mail: pantev@uni-muenster.de.

physiological evidence to the above conclusions and thus emphasize the significance of dichotic pitch phenomena.

A number of studies (Elberling et al., 1982; Romani et al., 1982; Pantev et al., 1988, 1994, 1995; Yamamoto et al., 1992) have shown that the functional magnetic imaging technique, based on magnetoencephalographic recordings (MEGs), is useful in mapping the tonotopy of the auditory cortex in humans. Furthermore, it has been shown that spatial location of the neural sources of the N1 component of the auditory evoked field (AEF) reflects the pitch rather than the spectral components of the stimulus (Pantev et al., 1989). This implies that tonotopic representation at the cortical level is fundamentally different from the tonotopic representation in the cochlea and the following tonotopically organised structures along the auditory pathway. Therefore, it is not possible to explain pitch perception solely on the basis of the response of auditory neurones to specific frequencies. Here we provide the first electrophysiological evidence from magnetic source imaging that a pitch processor fuses the information coming dichotically in order to produce a virtual pitch represented in the same cortical area as the spectral pitch.

2. Methods

In the present investigation, the auditory evoked magnetic responses to pure tones and complex tones were measured in two female and six male subjects aged between 19 and 36 years with no history of otological or neurological disorders and normal audiological status (air conduction and bone conduction thresholds no more than 10 dB hearing level in the range from 250 to 8000 Hz). All subjects were right-handed (Edinburgh handedness questionnaire, Oldfield, 1971). Informed consent was obtained from each subject after the nature of the study was explained to them. The subjects were paid for their participation.

Four different burst stimuli (duration 500 ms) were presented with a 10 ms cosine slope that was steep enough to elicit a prominent N1 component of the AEF, but gradual enough to keep a narrow frequency spectrum. The four stimuli differed in their spectral contents (Fig. 1) as follows: (1) a 250 Hz pure tone; (2) 1500 Hz pure tone; (3) a complex tone composed of eight components, these being the 4th to 11th harmonics of the missing fundamental frequency of 250 Hz (all presented monaurally to the left ear); and (4) two complex tones composed of the even and odd harmonics of the 8-component complex tone (presented dichotically, even harmonics presented to the left and odd harmonics to the right ear). Stimulus intensity was 60 dB relative to the individual hearing threshold. The interstimulus interval was randomised between 2.5 and 3.5 s. The four differ-

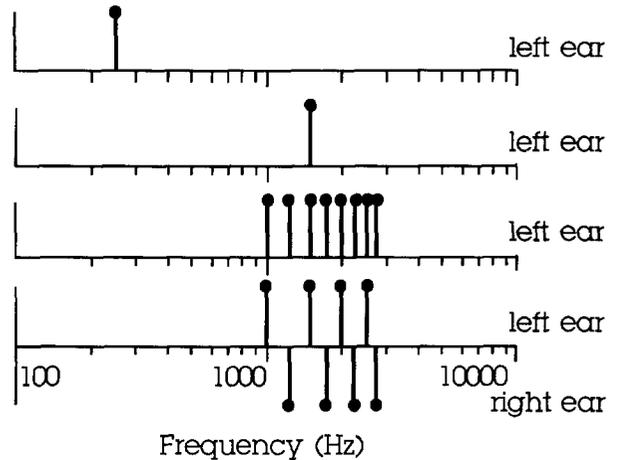


Fig. 1. Scheme of the spectral content of the stimuli. First and second rows: spectral pitches of 250 and 1500 Hz presented to the left ear (upward pointing pin). Third row: virtual pitch consisting of the 4th to 11th harmonics of the missing fundamental frequency of 250 Hz presented to the left ear (upward pointing pin). Fourth row: two complex tones composed of the even and the odd harmonics of the 8-component complex tone presented dichotically (upward and downward pointing pins).

ent stimuli were presented in 16 blocks of 150 identical stimuli each and the block sequence was varied across subjects in a randomised order. The magnetically silent delivery of the stimuli required a special delivery system with speakers (compressor driver type) outside the magnetically shielded room and echo-free tone transmission through 6.3 m of plastic tubing (16 mm inner diameter) to a silicon ear piece. This system was able to provide almost linear frequency characteristics in the range between 200 and 4000 Hz (deviations less than ± 4 dB, Pantev et al., 1991).

Auditory magnetic fields evoked by the different stimuli were recorded simultaneously from the left and the right hemisphere. Two magnetic sensor arrays (BTi-Magnes system), each containing 37 channels and covering a circular area with a diameter of 144 mm (first-order axial gradiometers with a baseline of 50 mm and coil diameter of 20 mm) were used to record the neuromagnetic brain activity simultaneously from both auditory cortices. The subject's head rested on the bottom sensor, supported by a soft pad. A sensor position indicator system determined the spatial locations of the sensors relative to the head and indicated if head movements occurred during the recordings. No head movements sufficient to require discarding of data were observed in the study. The sensor arrays were centred over T3 and T4 locations of the 10/20 system for electrode placement. Four blocks of 150 stimulus-related epochs each (bandwidth 1–100 Hz), belonging to the same stimulus condition, were sub-averaged and their consistency was checked. In the case of correspondence of the sub-averages (test and $3 \times$ re-test) they were cross-averaged. Thus the AEF that was used

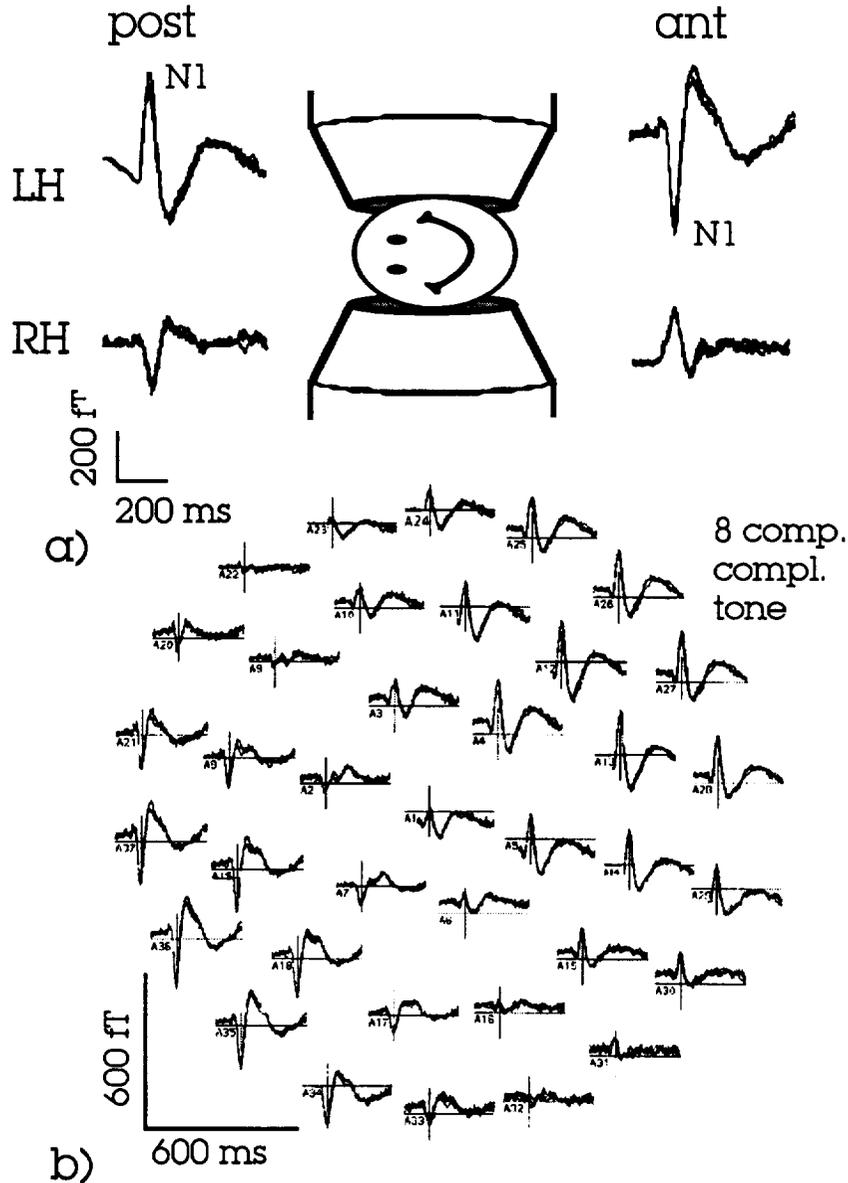


Fig. 2. (a) Simultaneous magnetoencephalographic recording of auditory evoked activity from the left and the right hemispheres. The waveforms at the posterior and anterior field extremes are superimposed for four repeated blocks of identical stimulation. (b) Evoked field distribution for the 8-component complex tone. The waveforms of the four repeated blocks ($N = 150$) are superimposed for comparison.

for the source analysis was an average of approximately 600 stimulus-related epochs (epochs contaminated by muscle or eye blink artefacts with signal variations of more than 3 pT were automatically rejected from the averaging procedure). In order to reduce possible variations caused by changes in attention and vigilance, subjects watched cartoons during the session.

For each AEF distribution corresponding to the four stimulus conditions a single equivalent current dipole model (ECD) located in a best fitting local sphere was used to estimate the moment and the space coordinates of a dipole, representing the cortical activity and the centre of the activated cortical area. The goodness of fit of the equivalent current dipole model indicates the percentage of the field variance that can be explained by

the model. The averaged value of the source coordinates of the data points around the N1 peak, having the highest 2% of goodness of fit was taken as a measure of the source locations for each of the four different stimulus conditions. These averaged N1 source locations were estimated in a head-based coordinate system. The origin of this coordinate system was set at the midpoint of the medio-lateral axis (y-axis) which joined the centre points of the entrance to the acoustic meatii of the left and the right ears (positive towards the left ear). The postero-anterior axis (x-axis) was oriented from the origin to the nasion (positive towards the nasion) and the inferior-superior axis (z-axis) was perpendicular to the x-y plane (positive towards the vertex). The estimated source locations were accepted

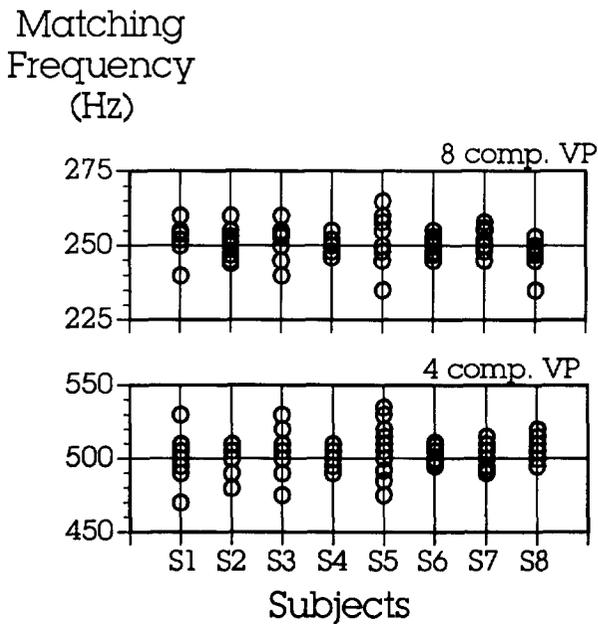


Fig. 3. Ten repeated matching results for each subject. The 8-component virtual pitch was matched with a spectral pitch of or close to 250 Hz, presented to the opposite ear (top row). The even or odd harmonics of the 8-component complex tone (periodicity of 500 Hz) were matched with a spectral pitch of or close to 500 Hz presented to the opposite ear (bottom row).

only in the case that their confidence volumes were smaller than 300 mm³. One-way analyses of variance (ANOVA) were calculated separately for the left and the right hemispheres, using Stimulus Class (the four different stimuli) as a repeated measures factor. Probabilities were Greenhouse–Geisser adjusted. Post-hoc Scheffé tests were used to resolve significant ANOVA effects.

The care of humans in this study were approved by the Ethics Commission of the Medical Faculty of the University of Münster and the MEG measurements were classified as completely non-invasive (permission of the Ethics Commission on Nov. 15, 1993).

3. Results

Fig. 2a demonstrates examples of averaged waveforms for the four blocks of identical stimuli, that were recorded from the anterior and posterior field maxima of the left and the right hemisphere. These four waveforms were superimposed for comparison. Realizing that about 25 min have passed between the collection of two successive blocks of epochs related to the same stimulus, the reliability of the neuromagnetic data obtained is remarkable. The different waveforms are almost identical in shape within the range of the N1 component. Due to the head-sensor constellation (cf. Fig. 2a) the sensing coils of the bottom sensor were further away from the scalp as compared with the

upper sensor. This fact causes a reduction of the AEF amplitudes recorded from the right hemisphere and consequently a worse signal-to-noise ratio. The well pronounced dipolarity of the field distribution can be easily recognized in Fig. 2b (in this case for the 8-component complex tone). This dipolar distribution was highly consistent through the different stimulus conditions and through the different test subjects and justifies the use of the ECD model for the experimental MEG data of this study.

The 8-component complex tone shows a 250 Hz periodicity and although it does not include spectral power below 1000 Hz, 250 Hz periodicity was audible. In 10 consecutive trials, all subjects were able to match this virtual pitch successfully with a spectral pitch of or close to 250 Hz, that was simultaneously presented to the opposite ear (Fig. 3, top). In the same way the complex tones containing the even or odd harmonics of the 8-component complex tone (periodicity of 500 Hz) were matched with a spectral pitch of or close to 500 Hz (Fig. 3, bottom). Since there was no substantial difference between the matching results to the odd and even harmonics of the 8-component complex tone, those results were presented together at the bottom part of Fig. 3.

The estimated cortical source locations, corresponding to the four different types of stimulus, are shown in Fig. 4 for the left hemisphere separately for each subject and also as spatial grand averages across subjects. Goodness of fit equalled 99% in the majority of the cases, confirming the adequacy of the ECD model. The spatial source coordinates were separately tested for stimulus-related differences.

The separation of the representation of the 1500 Hz source from a location around which the other three sources cluster was corroborated by statistical analysis. In the medio-lateral direction the effect was significant ($F(3, 21) = 16.39$; $\epsilon = 0.45$; $P < 0.003$). Post-hoc Scheffé testing revealed that this effect was exclusively due to the 1500 Hz spectral pitch condition which produced a location different from each of the other three conditions (all $P < 0.001$). On average, this difference amounted to 6.3 mm. Similar effect was also evident for the postero-anterior direction ($F(3, 21) = 9.35$; $\epsilon = 0.44$; $P < 0.02$). The 1500 Hz spectral pitch source was located 2.9 mm more posteriorly to the sources evoked by the other three stimulus conditions.

Since only data with a goodness of fit greater than 95% were accepted, only four subjects were suitable for the analyses of the right hemisphere. This fact can be explained with the worse signal-to-noise ratio, as demonstrated in Fig. 2a. Therefore, one-way analyses of variance (ANOVA) were calculated separately for the left and the right hemispheres, using Stimulus Class (the four different stimuli) as a repeated measures factor.

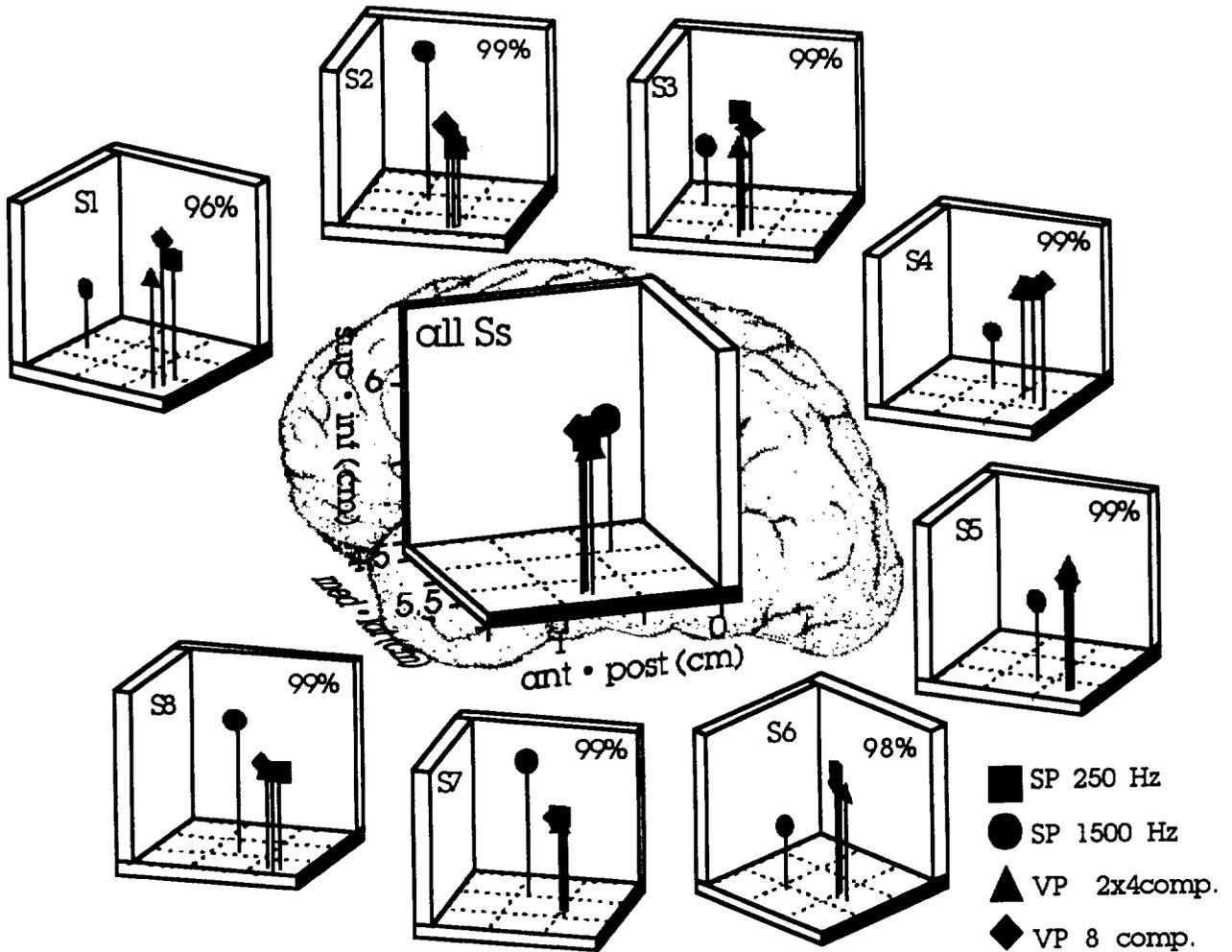


Fig. 4. Source locations of the activated cortical areas in the left hemisphere corresponding to the four different types of stimulation are presented individually for each subject around a sketched brain indicating the position. Subject's number and goodness of fit are listed. The enlarged view at the centre represents the spatial grand average across subjects. Squares denote spectral pitch of 250 Hz, circles spectral pitch of 1500 Hz, diamonds virtual pitch of the 8-component complex tone that were presented monaurally; triangles virtual pitch of the even and odd harmonics of the 8-component complex tone presented dichotically. The antero-posterior axis is denoted by a filled bar, the dimension of each cube is $1.5 \times 1.5 \times 1.5$ cm.

In general the results obtained from the right hemisphere are in line with those obtained from the left hemisphere. As demonstrated in the Fig. 5 the source locations for the pure tones of 1500 Hz were also clearly separated from those of the other three stimuli, which were closely clustered. This effect was found to be significant for the medio-lateral direction ($F(3, 9) = 31.62$; $\epsilon = 0.72$; $P < 0.001$) with a mean difference of 5.5 mm. However, in contrast to the left hemisphere, no significant difference in postero-anterior direction was found for the right hemisphere.

4. Discussion

If the psychoacoustic prediction of 'binaural fusion' is correct, then the activated brain regions corresponding to the 8-component virtual pitch, presented monau-

rally or dichotically, should activate the same cortical region as the one activated by a 250 Hz pure tone. If there were no binaural fusions, then the brain sources for the dichotically presented virtual pitch would be located consistently more medially than the representation of the 250 Hz tone due to their matching frequencies of 500 Hz (Pantev et al., 1990, 1994, 1995).

The finding that the same cortical region is activated by the 250 Hz spectral pitch and the 250 Hz virtual pitch (the complex tone composed of eight harmonics of the missing fundamental frequency of 250 Hz) confirms that the perceived pitch and not simply the spectral content of the auditory stimulus determines the representation within the tonotopic cortical map. This implies that the virtual pitch has been determined at a stage prior to the AEF N1 peak, i.e., the hypothesized pitch processor should be realized in a neuronal net-

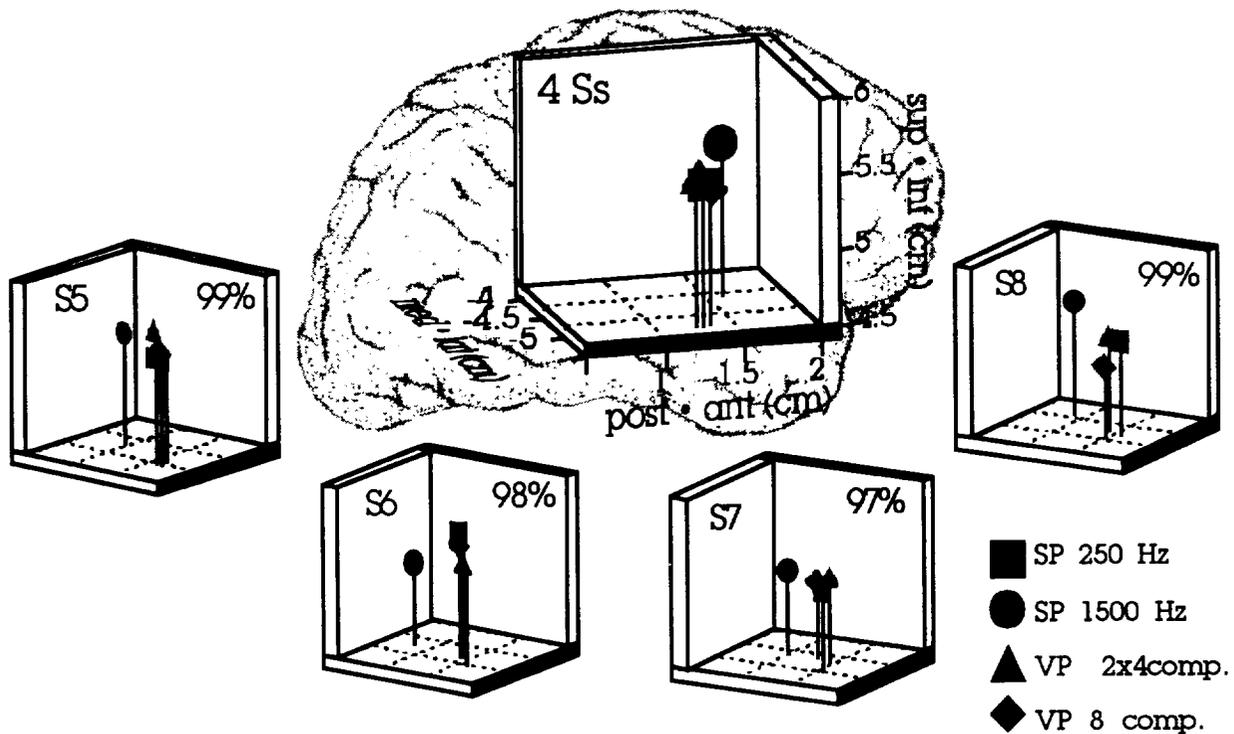


Fig. 5. Source locations of the activated cortical areas in the right hemisphere corresponding to the four different types of stimulation as obtained in four out of eight subjects. Subject's number and goodness of fit are listed. The enlarged view at the centre represents the spatial grand average across subjects. Squares denote spectral pitch of 250 Hz, circles spectral pitch of 1500 Hz, diamonds virtual pitch of the 8-component complex tone that were presented monaurally; triangles virtual pitch of the even and odd harmonics of the 8-component complex tone presented dichotically. The antero-posterior axis is denoted by a filled bar, the dimension of each cube is $1.5 \times 1.5 \times 1.5$ cm.

work of the primary auditory cortex or in sub-cortical structures. If there was no fusion of the complex tone harmonics presented dichotically, then the activated cortical area should reflect the 500 Hz region of the cortical tonotopic map. This region has been found to be several millimetres significantly more medial than the 250 Hz cortical area (Pantev et al., 1995). However, the statistical evaluation of our data clearly show that this is not the case. Given the absence of a difference in the location of the cortical excitation produced by the 8-component complex tone when presented either monaurally or dichotically, the results can be considered the first neurophysiological confirmation of binaural fusion in hearing. Furthermore, the data suggest that binaural fusion produces equal results in both hemispheres. Neural excitations travelling along auditory pathways are melded into one pattern at or before the cortical level. At the cortex, we observe the non-linear addition of information from both ears, having a quality that pertains to a virtual pitch composed of the dichotically presented complex tone. In conclusion, virtual pitch is not determined separately for the input to each ear, but is the result of binaural fusion in much the same way as the perception of monaurally presented complex tones. This discovery of neuromagnetic correlates of binaural fusion in hearing provides a new constraint for theories of pitch of complex tones.

Acknowledgments

This work was supported by grants from the Deutsche Forschungsgemeinschaft (Klinische Forschergruppe 'Biomagnetismus und Biosignalanalyse').

References

- Bilsen, F.A. (1976) Pronounced binaural pitch phenomenon. *J. Acoust. Soc. Am.* 59, 467–468.
- Bilsen, F.A. (1977) Pitch of noise signals: evidence for a 'central spectrum'. *J. Acoust. Soc. Am.* 61, 150–161.
- Bilsen, F.A. and Goldstein, J.L. (1974) Pitch of dichotically delayed noise and its possible spectral basis. *J. Acoust. Soc. Am.* 55, 292–296.
- Cramer, E.M. and Huggins, W.H. (1958) Creation of pitch through binaural interaction. *J. Acoust. Soc. Am.* 30, 413–417.
- Durlach, N.I. (1962) Note on the creation of pitch through binaural interaction. *J. Acoust. Soc. Am.* 34, 1096–1099.
- Elberling, C., Bak, C., Kofoed, B., Lebech, J. and Saermark, K. (1982) Auditory magnetic fields. Source localization and 'tonotopical organization' in the right hemisphere of the human brain. *Scand. Audiol.* 11, 61–65.
- Goldstein, J.L. (1973) An optimum processor theory for the central formation of the pitch of complex tones. *J. Acoust. Soc. Am.* 54, 1496–1516.
- Houtsma, A.J.M. (1979) Musical pitch of two-tone complexes and predictions by modern pitch theories. *J. Acoust. Soc. Am.* 66, 87–99.

- Houtsma, A.J.M. and Goldstein, J.L. (1972) The central origin of the pitch of complex tones: evidence from musical interval recognition. *J. Acoust. Soc. Am.* 51, 520–529.
- Klein, M.A. and Hartmann, W.M. (1981) Binaural edge pitch. *J. Acoust. Soc. Am.* 70, 51–61.
- Oldfield, R.C. (1971) The assessment and analysis of handedness: the Edinburgh Inventory. *Neurophysiologia* 9, 97–113.
- Pantev, C., Hoke, M., Lehnertz, K., Lütkenhöner, B., Anogianakis, G. and Wittkowski, W. (1988) Tonotopic organization of the human auditory cortex revealed by transient auditory evoked magnetic fields. *Electroenceph. Clin. Neurophysiol.* 69, 160–170.
- Pantev, C., Hoke, M., Lütkenhöner, B. and Lehnertz, K. (1989) Tonotopic organization of the auditory cortex: pitch versus frequency representation. *Science* 246, 486–488.
- Pantev, C., Hoke, M., Lehnertz, K., Lütkenhöner, B., Fahrendorf, G. and Stöber, U. (1990) Identification of sources of brain neuronal activity with high spatiotemporal resolution through combination of neuromagnetic source localization (NMSL) and magnetic resonance imaging (MRI). *Electroenceph. Clin. Neurophysiol.* 75, 173–184.
- Pantev, C., Gallen, C., Hampson, S., Buchanan, S. and Sobel, D. (1991) Reproducibility and validity of neuromagnetic source localization using a large array biomagnetometer. *Am. J. EEG Technol.* 31, 83–101.
- Pantev, C., Eulitz, C., Elbert, T. and Hoke, M. (1994) The auditory evoked sustained field: origin and frequency dependence. *Electroenceph. Clin. Neurophysiol.* 90, 82–90.
- Pantev, C., Bertrand, O., Eulitz, C., Verkindt, C., Hampson, S., Schuirer, G. and Elbert, T. (1995) Specific tonotopic organizations of different areas of the human auditory cortex revealed by simultaneous magnetic and electric recordings. *Electroenceph. Clin. Neurophysiol.* 94, 26–40.
- Romani, G.L., Williamson, S.J. and Kaufman, L. (1982) Tonotopic organization of the human auditory cortex. *Science* 216, 1339–1340.
- Terhardt, E. (1972) Zur Tonhöhenwahrnehmung von Klängen. I. Psychoakustische Grundlagen. *Akustica* 26, 173–186.
- Terhardt, E. (1987) Psychophysics of audio signal processing and the role of pitch in speech. In: M.E.H. Schouten (Ed.), *The Psychophysics of Pitch Perception*. Nijhoff, Dordrecht, pp. 271–283.
- Terhardt, E. (1989) Warum hören wir Sinustöne? *Naturwissenschaften* 76, 496–504.
- Van den Brink, G. (1974) Monotic and dichotic pitch matchings with complex sounds. In: E. Zwicker and E. Terhardt (Eds.), *Facts and Models in Hearing*. Springer, Heidelberg, pp. 178–189.
- Yamamoto, T., Uemura, T. and Llinás, R. (1992) Tonotopic organization of human auditory cortex revealed by multi-channel SQUID system. *Acta Otolaryngol. (Stockh.)* 112, 201–204.