Cortical processing of change detection: Dissociation between natural vowels and two-frequency complex tones

Minna Vihla*, Olli V. Lounasmaa, and Riitta Salmelin

Brain Research Unit, Low Temperature Laboratory, Helsinki University of Technology, Post Office Box 2200, 02015 HUT, Espoo, Finland

We compared magnetoencephalographic responses for natural vowels and for sounds consisting of two pure tones that represent the two lowest formant frequencies of these vowels. Our aim was to determine whether spectral changes in successive stimuli are detected differently for speech and nonspeech sounds. The stimuli were presented in four blocks applying an oddball paradigm (20% deviants, 80% standards): (*i*) /ɑ/ tokens as deviants vs. /i/ tokens as standards; (*ii*) /e/ vs. /i/; (*iii*) complex tones representing /ɑ/ formants vs. /i/ formants; and (*iv*) complex tones representing /e/ formants vs. /i/ formants. Mismatch fields (MMFs) were calculated by subtracting the source waveform produced by standards from that produced by deviants. As expected, MMF amplitudes for the complex tones reflected acoustic deviation: the amplitudes were stronger for the complex tones representing /ɑ/ than /e/ formants, i.e., when the spectral difference between standards and deviants was larger. In contrast, MMF amplitudes for the vowels were similar despite their different spectral composition, whereas the MMF onset time was longer for /e/ than for /ɑ/. Thus the degree of spectral difference between standards and deviants was reflected by the MMF amplitude for the nonspeech sounds and by the MMF latency for the vowels.

The ability to detect differences between sounds that differ in their spectral composition forms the basis for distinguishing phonemes and, ultimately, for understanding speech. The recognition of vowels requires that we can perceive the invariance caused by their formant structure despite intracategory variation of fundamental frequency (F0), sound duration, or intensity. Thus, to convey linguistic information, vowels must be perceived as representatives of certain phonological categories, whereas intracategory acoustic variation and the degree of intercategory acoustic deviation can be disregarded. The categorical perception of vowels has been explained by the perceptual magnet effect, which means that categorization diminishes the ability to detect within-category differences and which has been argued to mould phonetic perception in infancy (1–3).

In the present work, we tested whether spectral changes in successive stimuli are detected differently for speech and nonspeech sounds. We used natural vowels uttered by various speakers to study phonological categories that contain phonologically irrelevant variation, as in normal speech.

In two pilot studies, we compared magnetoencephalographic responses peaking ≈100 ms after the stimulus onset (N100m) (4) for natural vowels, presented in equal numbers, to detect possible temporal or spatial differences among them. In previous studies, the existence of clear location differences among vowels has been questioned, whereas N100m latency differences have been observed among the vowels /ɑ/, /i/, and /u/ (5, 6). Our pilot studies did not reveal location, orientation, N100m latency, or N100m peak amplitude differences between five Finnish vowels.

We therefore proceeded to our main experiment, an oddball paradigm, to detect mismatch fields (MMFs). Mismatch negativity and its magnetic counterpart MMF are considered to reflect the detection of dissimilarity between the signal and the stimulus represented by previous auditory memory traces (7).

Mismatch responses have been observed for rising vs. falling glides (8) and phonetic contrasts (9, 10). A phonetically relevant change in the stimuli (a transition of the second formant signaling a consonantal change) has been found to produce an MMF, despite variation of F0 in both standard and deviant stimuli (11).

We investigated, first, whether cortical change detection processes distinguish between /ɑ/ and /e/, each presented with the standard vowel /i/ and, second, whether the possible distinction is similar for vowels and complex tones representing the two lowest formant frequencies (F1 and F2) of these same vowels. If the acoustic distance between F1 and F2 is a decisive factor in the early cortical processing of vowels, similar spectral changes should produce similar MMF changes for vowels and respective two-tone composites.

Methods

Stimuli. *Pilot study 1.* Five natural Finnish vowels (/ɑ/, /e/, /i/, /o/, and /u/) were uttered by 20 Finnish female speakers (19–37 yr; students of speech pathology, one researcher). The vowels were recorded by using a digital audio tape recorder (DA-7 Casio, Dover, NJ), with the speaker's mouth at a distance of 30 cm from the microphone. The speakers were instructed to utter the vowels by using constant height, length, and loudness. Various samples of each vowel were recorded, and perceptually good representatives were selected for the analysis, 100 tokens in all. Table 1 shows the mean (± SD) F0 and duration of the samples. The vowels were edited to have the same peak amplitude (as measured using the Macromedia SOUNDEDIT program). F0 or sound duration were not edited, to keep the perceived quality of the vowels unaltered. The variation of F0 and duration was similar in the different vowel categories.

The stimuli were presented to the subject's right ear at 70 dB above hearing threshold in a quasi-random order such that two tokens of the same vowel or two samples by the same speaker never occurred in a row. The hearing threshold was defined by using a 1-kHz 50-ms tone. The onset-to-onset sound interval was 2 s.

Pilot study 2. From the stimuli used in pilot study 1, a reduced set of stimuli of exceptionally high perceptual quality was selected, including three vowels (/ɑ/, /e/, and /i/) uttered by six of the speakers (24–37 yr). The Finnish vowel system includes eight vowels: five front vowels, two of which are rounded (/i/ as in sit, /y/ as in French duc, /e/ as in set, /œ/ as the eu-sound in French

---

Table 1. Mean F0 and duration of Finnish vowels, spoken by females

Vowel	Mean ± SD	
	F0, Hz	Duration, ms
	Twenty speakers	
/ɑ/	203 ± 19	134 ± 27
/e/	205 ± 29	135 ± 29
/i/	201 ± 20	132 ± 28
/o/	193 ± 18	122 ± 24
/u/	200 ± 24	118 ± 23
	Six speakers	
/ɑ/	189 ± 12	139 ± 20
/e/	187 ± 15	139 ± 25
/i/	183 ± 15	145 ± 36

F0 was measured manually from the signal waveform. Samples by 20 speakers were used in the first pilot study, and samples by 6 speakers in the second pilot study and in the main study.

leur, /æ/ as in hat; /y/ and /œ/ are rounded), and three back vowels (/u/ as in put, /o/ as in top, and /ɑ/ as in but). We selected the reduced set of vowels so that it included a pair that represents opposite corners within the vowel space (/i/ and /ɑ/) as well as a pair that is more similar acoustically (/i/ and /e/).

We compared responses to /i/ in three different conditions: (*i*) /i/ presented alone; (*ii*) /i/ and /e/ randomized, equiprobable; and (*iii*) /i/ and /ɑ/ randomized, equiprobable, with an onset-to-onset interval of 1.3 s. Two samples by the same speaker never occurred in a row. The order of the blocks varied across the subjects. The stimuli were presented to the right ear at 70 dB HL, with the hearing threshold determined as in pilot study 1.

Main experiment. The stimuli included natural vowels and two-frequency complex tones. The vowels were the same as in pilot study 2. The vowel samples of each speaker were analyzed by using fast Fourier transform (the MathWorks MATLAB program), and the frequency peaks that represent F1 and F2 were estimated with an accuracy of 50 Hz. Fig. 1 shows the formant frequencies; they were used as the frequency components of the complex tones. Because the F1 and F2 values of a certain vowel remained invariant across some of the speakers, the total number of different F1–F2 combinations, and thus the total number of the complex tones, was 11, instead of the 18 vowel samples (three vowels by six speakers). The duration of the complex tones was 150 ms, including gradual 75-ms fading-in and 75-ms fading-out
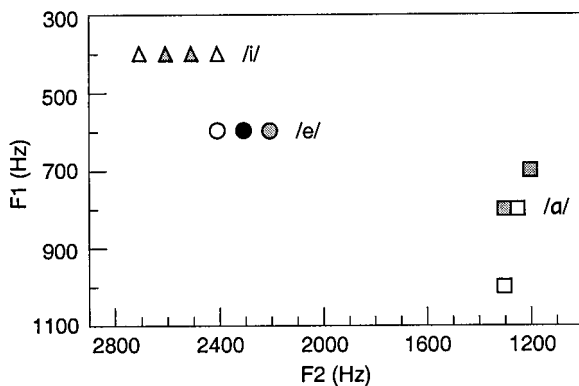
times. The complex tones were edited to have the same peak amplitude as the vowels.

The stimuli were presented in four blocks applying an oddball paradigm (80% standards, 20% deviants): (*i*) /ɑ/ tokens as deviants vs. /i/ tokens as standards; (*ii*) /e/ tokens as deviants vs. /i/ tokens as standards; (*iii*) complex tones representing F1 and F2 of the /ɑ/ tokens as deviants vs. complex tones representing the /i/ tokens as standards; and (*iv*) complex tones representing the /e/ tokens as deviants vs. complex tones representing the /i/ tokens as standards.

The onset-to-onset interval between the stimuli was 1 s. Two deviant stimuli never occurred in a row, and, with the vowels, each stimulus was followed by a vowel uttered by a different speaker. The complex tones representing formant frequencies were presented in the same order as the respective vowels. The order of the blocks was counter-balanced across subjects. The recording of one block lasted 15 min, and there was a brief pause between successive blocks, during which the subject was allowed to adjust his position. The subjects were silently reading a book to keep their attention level stable.

The stimuli were presented to the right ear at 70 dB HL, as measured in the pilot studies. The hearing threshold also was measured by using the vowel /ɑ/ and complex tones, which gave a slightly higher threshold than the measurement using the 1-kHz tone. However, when the playback intensity was adjusted to the hearing threshold as defined by using the 1-kHz tone, the stimuli were loud enough to be heard clearly and yet not disturbing.

Subjects. The subjects were Finnish-speaking, right-handed volunteers without known neurological or hearing problems. They gave their informed consent to participate in this study. Three subjects (25–33 yr, one male) participated in pilot study 1, four female subjects in pilot study 2 (25–38 yr), and eight male subjects in the main experiment (23–30 yr).

Recording. Magnetoencephalographic signals were recorded in a magnetically shielded room, using a helmet-shaped 306-channel whole-head neuromagnetometer (Vectorview, Neuromag, Helsinki, Finland). It contains triple sensor elements at 102 locations, 306 superconducting quantum interference detectors in all. Each sensor element consists of two orthogonally oriented planar gradiometers, which detect maximum signal directly above the active cortical area, and one magnetometer. In the present study, we used only data from the planar gradiometers. The passband was 0.03–200 Hz, and the sampling rate was 600 Hz. Vertical and horizontal electro-oculograms were recorded for on-line rejection of epochs contaminated by excessive eye movement and blinking artifacts. The magnetoencephalographic responses were averaged over a 700-ms interval, including a 200-ms prestimulus baseline. A minimum of 150 responses to each vowel was collected in the pilot studies and a minimum of 120 responses to deviants in the main experiment.

The location of the subject's head with respect to the device was determined by using four coils attached to the subject's head. First, the coils were located within the head coordinate system defined by three anatomical landmarks (preauricular points and nasion), using a three-dimensional digitizer. Second, the coils were located within the magnetoencephalographic device coordinate system by energizing them briefly before each recording session. The location of the active areas could thus be expressed in head coordinates and presented on structural MRIs of individual subjects.

Data Analysis. Data were low-pass filtered at 40 Hz before source analysis. The responses were modeled by using equivalent current dipoles (ECD), which estimate the mean location, orientation, and strength of the cortical current flow from the distribution of the magnetic field (12). In all experiments and for all

Fig. 1. The first two formant frequencies of the vowels /ɑ/, /e/, and /i/, spoken by six Finnish female speakers. The dots represent individual samples. The 18 vowel samples are represented by 11 F1–F2 combinations, due to close similarities in the spectra of individual speakers. The color of the dots indicates how many samples they represent: white, one speaker; gray, two speakers; and black, three speakers.

stimuli, activity in the time interval 50–160 ms after the stimulus onset was adequately accounted for by two stable dipoles, one in the auditory cortex of each hemisphere. ECDs were determined, separately in each hemisphere, on the basis of ≈20 sensor pairs during the most prominent peak (N100m). Thereafter, the locations and orientations of the two ECDs were kept fixed, while their amplitudes were allowed to vary to best explain the field pattern detected by all sensors during the whole recording period.

For every subject, ECDs were first determined separately for each stimulus. The pilot studies with three and four subjects did not show any obvious differences in ECD parameters. In the main experiment, ECD locations and orientations were tested for systematic effects of hemisphere, stimulus type (vowel vs. complex tone), and stimulus quality (vowels/complex tones representing /ɑ/, /i/ with /ɑ/, /e/, or /i/ with /e/, or respective formant frequencies) using a $2 \times 2 \times 4$ repeated measures ANOVA. The only significant difference among the stimuli was that the current flow, approximately perpendicular to the course of the Sylvian fissure, was oriented in the sagittal plane on average 9° more horizontally for the complex tones than the vowels, similarly in both hemispheres [$F(1,7) = 8.617$, $P < 0.03$].

As this small difference in orientation was the only difference between the stimuli, all of each subject's responses could be modeled with a single pair of dipoles to allow systematic comparisons of the different conditions. For each individual, a representative set of two ECDs was selected, with the clearest field pattern in the sensors over the left and right temporal areas (goodness-of-fit values between 88% and 97%). This two-dipole model explained typically 85–90% of the activity detected by all sensors during the N100m peak in all subjects and all stimulus conditions.

MMFs were calculated by subtracting the source waveform produced by standards from that produced by deviants. The effects of hemisphere, stimulus type (vowel vs. complex tone), and deviant (/ɑ/ or its formant frequencies vs. /e/ or its formant frequencies) on MMF onset latencies, peak latencies, and peak amplitudes were tested using a $2 \times 2 \times 2$ repeated measures ANOVA. The MMF onset time was defined as the point when the MMF amplitude curve rises above the baseline level immediately before its peak.

Results

The strongest responses occurred ≈100 ms after the stimulus onset over the bilateral auditory cortices. In the main experiment, the mean latency of the N100m peak was 98 ± 15 ms in the left hemisphere (LH), contralateral to the stimulated ear, and 109 ± 15 ms in the ipsilateral right hemisphere (RH). A later, more anterior response was seen 200 ms after the stimulus onset, although this field was too weak to be included in source modeling. The MMF overlapped with the N100m response, peaking at 131 ± 21 ms in LH and at 139 ± 21 ms in RH. The location of the N100m sources (determined separately for each stimulus condition) was on average 7 mm more anterior in RH than LH [$F(1,7) = 18.365$, $P < 0.01$], in agreement with previous reports (13–15).

Fig. 2 depicts one subject's dipole locations in the bilateral auditory cortices, ECD waveforms for deviants and standards, and difference waveforms indicating MMFs. The MMF overlaps the N100m in all stimulus conditions, with the exception of /e/ as the deviant, for which the MMF is delayed. For the complex tones representing /ɑ/ formants, the MMF is particularly prominent in RH.

Fig. 3 shows the mean ± SEM MMF onset latencies, peak latencies, and peak amplitudes across all subjects. In one subject, /e/ did not produce a detectable MMF, and thus data concerning MMF latencies are based on seven subjects. The MMF onset latencies were on average 19 ms longer for /e/ than /ɑ/, whereas
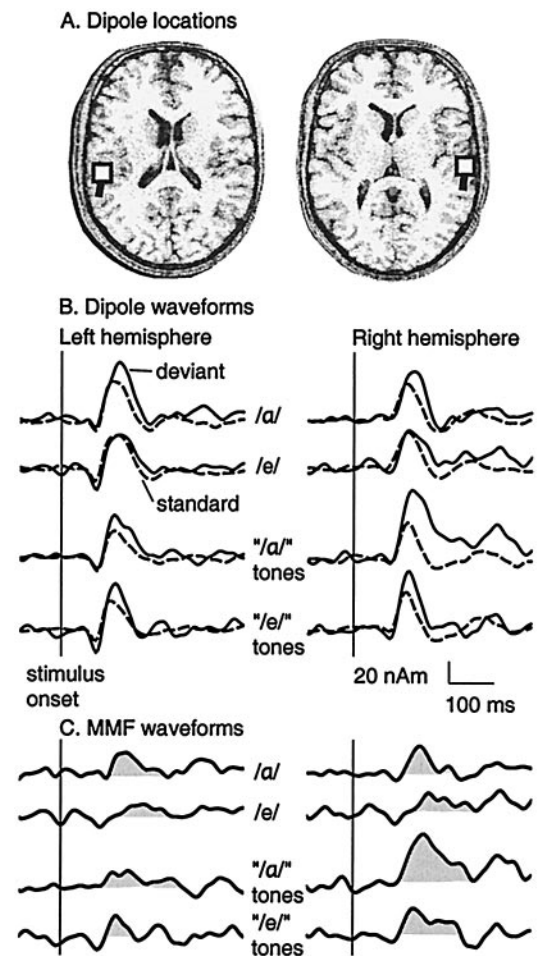

A. Dipole locations

B. Dipole waveforms

Left hemisphere   Right hemisphere

— deviant
/ɑ/
/e/
standard
"/ɑ/" tones
"/e/" tones
stimulus onset
20 nAm
100 ms

C. MMF waveforms
/ɑ/
/e/
"/ɑ/" tones
"/e/" tones

Fig. 2. One subject's responses for vowels and for complex tones consisting of the first two formant frequencies of the vowels. (*A*) Dipole locations (squares) in the left and right hemisphere. (*B*) Dipole waveforms for deviants (20%, solid line) and standards (80%, dashed line). The stimuli were presented in four blocks: *i*) the vowel /ɑ/ as the deviant vs. the vowel /i/ as the standard; *ii*) /e/ vs. /i/; *iii*) complex tones representing /ɑ/ formants ("/ɑ/" tones) vs. complex tones representing /i/ formants; and *iv*) complex tones representing /e/ formants ("/e/" tones) vs. complex tones representing /i/ formants. (*C*) MMFs were calculated by subtracting the source waveform for the standard from that for the deviant.

no differences were evident for the complex tones, as indicated by the significant stimulus type-by-deviant interaction [$F(1,6) = 23.033$, $P < 0.01$]. The MMF peak latencies were on average 15 ms longer for /e/ than /ɑ/, although the difference did not reach statistical significance [type-by-deviant interaction $F(1,6) = 4.922$, $P < 0.07$]. When the vowels were tested separately, using a hemisphere-by-deviant 2×2 repeated measures ANOVA, the MMF peak latencies were significantly longer for /e/ than /ɑ/ [$F(1,6) = 22.9$, $P < 0.01$].

The MMF amplitude for the complex tones representing /ɑ/ formants was on average 1.7 times the amplitude for the complex tones representing /e/ formants, whereas no amplitude differences were found among the vowels [type-by-deviant interaction $F(1,7) = 19.181$, $P < 0.01$]. The amplitudes were stronger for the complex tones than the vowels [$F(1,7) = 20.525$, $P < 0.01$]. As Fig. 3 shows, the amplitude difference between the complex tones and the vowels was more pronounced in RH than LH, although the hemisphere-by-type interaction did not quite reach significance [$F(1,7) = 5.434$, $P = 0.053$].
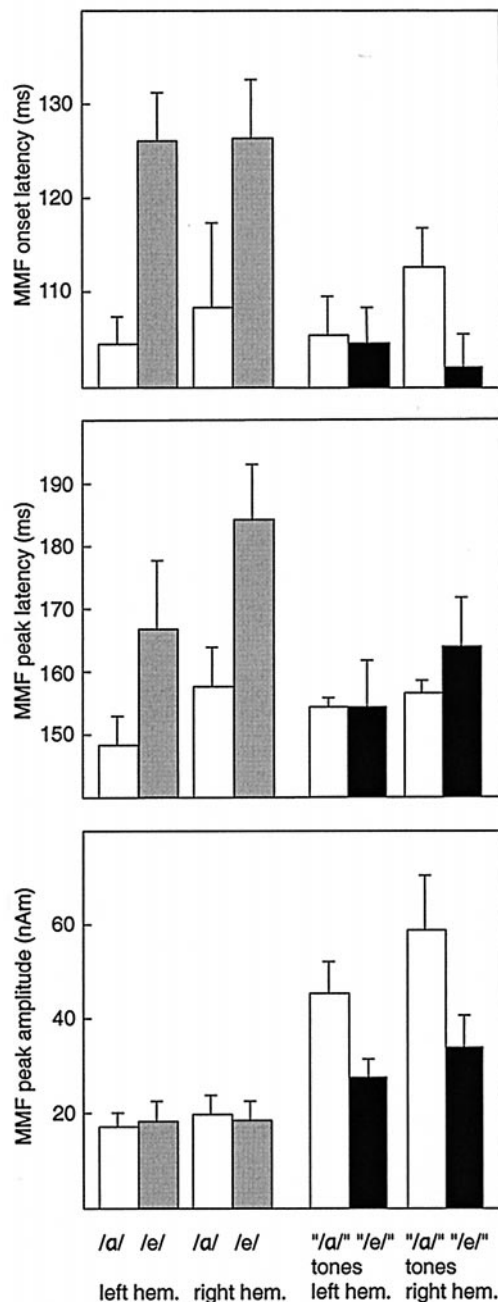
Vihla *et al.*

Fig. 3. Mean (± SEM) MMF onset latencies, peak latencies, and peak amplitudes for the vowels /ɑ/ and /e/ (white and gray bars) and complex tones consisting of the first two formant frequencies of the vowels ("/ɑ/" and "/e/" tones; white and black bars). The experimental paradigm is explained in the legend of Fig. 2.

Discussion

We compared responses to Finnish vowels and to nonspeech sounds that consisted of two tones representing the two lowest formant frequencies of the vowels. The N100m was the most prominent response, and the poststimulus activity was satisfactorily explained by two dipoles, one in each auditory cortex.

Dissociation of Change Detection Processes for Vowels and Nonspeech Sounds. The complex tones differed in their MMF amplitudes, consistent with previous findings suggesting that greater frequency deviation enhances the mismatch negativity elicited by simple tones (16, 17). In contrast, the acoustic difference between the vowels /ɑ/ and /e/, when compared with /i/, was reflected by their different MMF latencies, not amplitudes. A decrease in the acoustic difference between standard and deviant vowels increased the MMF latency, as has been previously reported for vowels (17) and consonant-vowel syllables (10). It is possible that a smaller acoustic difference between standard and deviant vowels in /e/ vs. /i/ than /ɑ/ vs. /i/ increases the time needed for the detection of dissimilarity. When the difference has been established, the MMF is equally strong for /e/ and /ɑ/.

A plausible explanation for our findings is that the MMF for vowels displays categorical (1–3) rather than acoustic discrimination. At the phonological level, /e/ and /ɑ/ are equally different from /i/. In addition to speech sounds, categorization training has been reported to decrease sensitivity for within-category acoustic differences by using nonspeech stimuli (18). The assumption that the MMF amplitude for vowels reflects categorical discrimination, rather than the degree of intercategory acoustic deviation, is in line with the observation that vowels that are phonemes in the subject's native language produce MMFs of roughly similar amplitude (17).

Stronger MMF for Nonspeech Sounds Than Natural Vowels. The MMF amplitudes were stronger for the complex tones than the vowels. A possible explanation for this MMF amplitude difference is the larger intracategory variation of the vowels. We used natural vowels spoken by various speakers, and so the tokens included in one vowel category differed from each other with respect to F0, duration, and the slope of amplitude rise and fall in the beginning and end of the stimuli. The vowels were edited as little as possible (only with respect to their peak amplitude) to preserve their original quality, and vowels uttered by various speakers were included because we wanted to compare types (phonological classes) rather than tokens (single representatives) of vowels. Unlike the vowels, the complex tones differed only with respect to their spectral components.

Previous studies suggest that MMFs for different features of auditory stimuli are additive (19) and that MMFs for frequency changes are weaker when other stimulus features (intensity, duration, envelope function, and harmonic structure) vary than when they are kept constant, possibly because the variation of other stimulus features either weakens the neural representation of frequency or dampens the mismatch process elicited by frequency changes (20). The MMF amplitude also is affected by interstimulus and interdeviant intervals (21); in the present work, the interstimulus interval varied slightly in the vowel blocks but not in the complex tone blocks, due to duration differences between the vowel samples. The stimulus onset asynchrony was, however, constant.

Although the mismatch response may be modulated by attention (22, 23), it is unlikely that the subjects would have paid more attention to the nonspeech than speech stimuli. In addition, possible shifts in attention do not explain why MMF amplitude differences were observed between the complex tones but not between the vowels.

Conclusions

Our findings suggest that changes in the spectral composition of successive stimuli are encoded differently for natural vowels and two-frequency complex tones. The vowels differed with respect to MMF timing, whereas the complex tones differed with respect to MMF amplitude. The lack of MMF amplitude differences between the vowels possibly reflects categorical discrimination.

NEUROBIOLOGY

1. Kuhl, P. K. (1991) *Percept. Psychophys.* **50,** 93–107.
2. Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N. & Lindblom, B. (1992) *Science* **255,** 606–608.
3. Kuhl, P. K. (1994) *Curr. Opin. Neurobiol.* **4,** 812–822.
4. Hari, R. (1990) in *Auditory Evoked Magnetic Fields and Electric Potentials, Advances in Audiology*, eds. Grandori, F., Hoke, M. & Romani, G. L. (Karger, Basel), Vol. 6, pp. 222–282.
5. Diesch, E., Eulitz, C., Hampson, S. & Ross, B. (1996) *Brain Lang.* **53,** 143–168.
6. Poeppel, D., Phillips, C., Yellin, E., Rowley, H. A., Roberts, T. P. L. & Marantz, A. (1997) *Neurosci. Lett.* **221,** 145–148.
7. Näätänen, R. (1992) *Attention and Brain Function* (Erlbaum, Hillsdale, NJ).
8. Pardo, P. J. & Sams, M. (1993) *Neurosci. Lett.* **159,** 43–45.
9. Diesch, E. & Luce, T. (1997) *Exp. Brain Res.* **116,** 139–152.
10. Alho, K., Connolly, J. F., Cheour, M., Lehtokoski, A., Huotilainen, M., Virtanen, J., Aulanko, R. & Ilmoniemi, R. J. (1998) *Neurosci. Lett.* **258,** 9–12.
11. Aulanko, R., Hari, R., Lounasmaa, O. V., Näätänen, R. & Sams, M. (1993) *NeuroReport* **4,** 1356–1358.
12. Hämäläinen, M., Hari, R., Ilmoniemi, R. J., Knuutila, J. & Lounasmaa, O. V. (1993) *Rev. Mod. Phys.* **65,** 413–497.
13. Kaukoranta, E., Hari, R. & Lounasmaa, O. V. (1987) *Exp. Brain Res.* **69,** 19–23.
14. Eulitz, C., Diesch, E., Pantev C., Hampson, S. & Elbert, T. (1995) *J. Neurosci.* **15,** 2748–2755.
15. Salmelin, R., Schnitzler, A., Parkkonen, L., Biermann, K., Helenius, P., Kiviniemi, K., Kuukka, K., Schmitz, F. & Freund, H.-J. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 10460–10465.
16. Tiitinen, H., May, P., Reinikainen, K. & Näätänen, R. (1994) *Nature (London)* **372,** 90–92.
17. Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, R. J., Luuk, A., *et al.* (1997) *Nature (London)* **385,** 432–434.
18. Guenther, F. H., Husain, F. T., Cohen, M. A. & Shinn-Cunningham, B. G. (1999) *J. Acoust. Soc. Am.* **106,** 2900–2912.
19. Levänen, S., Hari, R., McEvoy, L. & Sams, M. (1993) *Exp. Brain Res.* **97,** 177–183.
20. Huotilainen, M., Ilmoniemi, R. J., Lavikainen, J., Tiitinen, H., Alho, K., Sinkkonen, J., Knuutila, J. & Näätänen, R. (1993) *NeuroReport* **4,** 1279–1281.
21. Imada, T., Hari, R., Loveless, N., McEvoy, L. & Sams, M. (1993) *Electroencephalogr. Clin. Neurophysiol.* **87,** 144–153.
22. Näätänen, R., Paavilainen, P., Tiitinen, H., Jiang, D. & Alho, K. (1993) *Psychophysiology* **30,** 436–450.
23. Alain, C. & Woods, D. L. (1997) *Psychophysiology* **34,** 534–546.