Early Parallel Processing of Auditory Word and Voice Information

Thomas R. Knösche, Sonja Lattner, Burkhard Maess, Michael Schauer, and Angela D. Friederici

Max Planck Institute of Cognitive Neuroscience, Leipzig, Germany

The present study investigates the relationship of linguistic (phonetic) and extralinguistic (voice) information in preattentive auditory processing. We provide neurophysiological data, which show for the first time that both kinds of information are processed in parallel at an early preattentive stage. In order to establish the temporal and spatial organization of the underlying neuronal processes, we studied the conjunction of voice and word deviations in a mismatch negativity experiment, whereby the listener's brain responses were collected using magnetoencephalography. The stimuli consisted of single spoken words, whereby the deviants manifested a change of the word, of the voice, or both word and voice simultaneously (combined). First, we identified the N100m (overlain by mismatch field, MMF) and localized its generators, analyzing N100m/MMF latency, dipole localization, and dipole strength. While the responses evoked by deviant stimuli were more anterior than the standard, localization differences between the deviants could not be shown. The dipole strength was larger for deviants than the standard stimulus, but again, no differences between the deviants could be established. There was no difference in the hemispheric lateralization of the responses. However, a difference between the deviants was observed in the latencies. The N100m/MMF revealed a significantly shorter and less variant latency for the combined stimulus compared to all other experimental conditions. The data suggest an integral parallel processing model, which describes the early extraction of phonetic and voice information from the speech signal as parallel and contingent processes. © 2002 Elsevier Science (USA)

INTRODUCTION

Recent research has shown that many aspects of human auditory input are analyzed by the auditory cortices within 200 ms after stimulus. Näätänen *et al.* (2001) described this phenomenon as "primitive intelligence in the auditory cortex." A principal means of investigating these capabilities has been the use of unattended repetitive stimuli intermixed with some rare and random deviants (oddball design).

Any discriminable change in some repetitive aspect of auditory input causes an electrophysiological response of the brain called *mismatch negativity* (MMN, Näätänen *et al.,* 1978; for a detailed review see Näätänen, 2001). The MMN is manifested in a frontocentral negativity between 100 and 200 ms after stimulus onset, as has been shown by a vast number of neurophysiological studies (see Näätänen, 2001; Näätänen *et al.,* 2001). The MMN effect is preattentive (Sams *et al.,* 1985; Näätänen *et al.,* 1993) and mainly localized bilaterally in the auditory cortex (Hari *et al.,* 1984; Giard *et al.,* 1995; Alho, 1995). It has also been demonstrated that the exact locus of the generators may depend on the dimension of the change (Giard *et al.,* 1995).

However, the MMN reflects not only acoustic features of the auditory input but also higher cognitive properties. For example, it has been shown that the MMN amplitude is larger if the mismatching speech sound is a prototypical phoneme in the subject's native language but is smaller in response to normative speech sounds (Näätänen *et al.,* 1997). Speech, however, contains many different types of information that are processed concurrently and in relation to each other, i.e., not only as phonemic, lexical, grammatical, and prosodic information, but also as clues of the speaker's mood, age, or sex.

A particularly interesting field of investigation is the functional dichotomy of the speech content and its form: On the one hand, there are acoustic cues which lead to a phonetic identification of speech sounds and, in consequence, of words as well as larger linguistic units. Here we term this kind of information *word information* (cf. Mullenix and Pisoni, 1990). Second, there is also extralinguistic information in the acoustic signal that conveys information about the speaker. This kind of information will be termed *voice information*. Both word and voice information are extracted from the speech signal, and experiments investigating either phonological processing or voice discrimination have shown that both are processed preattentively (Alho *et al.,* 1998; Titova and Näätänen, 2001).

The present study is concerned with the localization and the temporal structure of the processes involved in word and voice deviance perception. To date, whether these deviations are processed by the same neural substrate and whether both hemispheres are equally involved in these processes remain unclear. There is no MMN study that makes a claim on the lateralization of the processing of voice deviancy. However, studies of phonagnostic patients have shown that voice discrimination abilities can be impaired by unilateral damage of either hemisphere. On the other hand, the ability to recognize familiar voices is clearly correlated to damage of the right hemisphere (Van Lancker and Kreiman, 1987; Van Lancker et al., 1988, 1989).

Functional imaging studies comparing human voices contrasted with nonvocal sounds report a bilateral activation of the superior temporal sulcus and superior temporal gyrus (Belin et al., 2000; Binder et al., 2000).[1] The only event-related potentials (ERP) study that directly investigates voice deviancy unfortunately reveals no information on the hemispherical distribution of the MMN (Titova and Näätänen, 2001). Another study by Levy and colleagues (2001) contrasting human voices with acoustically matched musical instruments makes no claims on MMN at all but comments only on the later P300 component. Those MMN studies investigating phonemic deviancy are not equivocal. Kasai et al. (2001) report a right lateralization for deviating vowels, if they belong to the same category as the standard. However, in the case of an across-category deviancy, the left hemisphere was dominant. On the other hand, there are many studies using consonant–vowel syllables, which find a left-dominant mismatch response (Shtyrov et al., 1998, 2000; Alho et al., 1998; Rinne et al., 1999). Finally, Mathiak et al. (2000) specify a bilateral involvement of the superior part of the temporal lobe during passive listening and left hemisphere dominance in a phoneme discrimination task.

Another important question for the present study concerns the relation between word and voice information in the processing of stimuli that are deviant in both dimensions simultaneously. One could assume a *serial model,* postulating that a certain type of information (e.g., word information) is checked for deviancy first, prior to another (e.g., voice information). Behavioral studies, however, point toward parallel processing of voice and word information. Reaction time experiments report interference in a "same–different" judgment of words, when word and voice information are varied randomly between trials (speeded-classification paradigm, Garner, 1974). In contrast, when fixed combinations of word and voice have been presented, a redundancy gain (shorter reaction time and

less variance in response latencies) has been observed (Wood, 1974; Mullenix and Pisoni, 1990).

The neural basis for these behavioral effects has not been established yet. MMN studies examining the multidimensional deviances in the domain of sinusoidal and complex tones, however, do indicate parallel processing (Winkler et al., 1992; Liasis et al., 2000). It is, however, unclear whether such a parallel processing mechanism is also present if the deviation comprises two cognitively higher aspects of the incoming speech sounds, namely word and voice information. In any case, such a mechanism would be much more economic than a serial one, especially if one considers the fact that there is a vast number of stimulus features that can elicit a mismatch response and the serial processing of which would result in a considerable delay (see, e.g., Näätänen et al., 2001).

If one assumes parallel processing of different kinds of deviant information, the question arises of whether each of these processes performs deviancy detection independently or in a contingent way. In the case of an independent manner of processing, the measured mismatch responses would simply add up; i.e., the amplitude of the MMN elicited by a feature conjunction would be the sum of the amplitudes of the MMNs caused by the corresponding single features. Note that this additivity assumption only holds if the MMN generators associated with the different stimulus features are located at cortical sites that are not very different in position and (even more importantly) orientation. This seems quite likely if MMN generators are located in the auditory cortex, as found by, e.g., Alho (1995). For some quite basic properties of auditory stimuli (stimulus onset asynchrony (SOA), frequency, duration, intensity), a certain degree of additivity of the MMN elicited by feature conjunctions has indeed been observed (Levänen et al., 1993; Wolff and Schröger, 2001; Paavilainen et al., 2001). However, additivity has not always been found, especially if three features have been altered at once or in the case of certain feature combinations (frequency and intensity) (Wolff and Schröger, 2001; Paavilainen et al., 2001). In these cases, the conjunction MMN was subadditive, i.e., its amplitude was less than the sum of the single-feature MMNs. On the other hand, Levänen and colleagues (1993) did demonstrate convincing additivity for the feature combinations of interstimulus interval + frequency and duration + frequency.

More specifically, one could assume an *integral model of parallel processing,* where different parallel processes gather "evidence" for the deviancy of the stimulus until a threshold is reached and the mismatch response is elicited. In such a model, the different deviancy detection processes do not act independently but contribute to one common mismatch detection process. This predicts that the amplitude of the MMN should not be affected by feature conjunc-

---

[1] Note that in an MEG study only sulcal activity can be reliably detected, since MEG is silent to currents directed toward or away from the head surface.

tions, but instead the speed and reliability of detection, e.g., the latency of the MMN.

Answering these questions concerning early auditory processing requires a precise spatiotemporal characterization of the generators underlying the early processing of word and voice information. Magnetoencephalography (MEG) has been shown to accurately localize the generators of early auditory brain responses (Elberling *et al.,* 1980; Hari *et al.,* 1980; Pantev *et al.,* 1988), including mismatch negativity (Levänen *et al.,* 1996; Yvert *et al.,* 2001).

In the present study we used an auditory mismatch paradigm with monosyllabic German words to investigate the effect of deviating word and voice information, presented both separately and in combination. MEG was employed to record the brain responses and dipole modeling served to reveal the spatiotemporal structure of the underlying generators. We characterized responses regarding (1) latency, (2) source location, and (3) source magnitude in order to decide between the various serial and parallel models for the processing of word and voice information. Furthermore, we predicted a mismatch effect of magnitude and source localization for all deviancy conditions. At least for word deviance, this effect might be stronger in the left hemisphere in accord with results reported by others regarding consonant–vowel syllables. According to the assumption that in the case of combined stimuli both processes work in parallel, we expected for this condition a shorter latency with less variance than for the pure word and voice conditions; in the case of serial processes, we predicted the reverse pattern. A precise mathematical modeling will be used to decide between models assuming *independent or dependent* integral parallel processing.

MATERIALS AND METHODS

Material and Experimental Design

Twenty healthy right-handed German-speaking volunteers participated in this study, 19 of which were successfully recorded, and 11 subjects (9 male) showed a clear response pattern and finally underwent the entire evaluation procedure (see below). All subjects reported normal hearing, gave their informed consent, and were paid for their participation. For each participant, the hearing threshold was determined and the stimulus presentation volume was fixed to 40 dB above this level.

An oddball design was employed. The German word "Test" ([tEst], test) was presented as the standard stimulus uttered by a male speaker, quasirandomly followed by one of three possible deviants:

• deviating word (WORD) uttered by the same male speaker (Dach/[dax]/roof),

TABLE 1

Overview of the Acoustic Properties of the Stimuli

	VOT	F0	F1	F2	F3
STANDARD /test/	55	82	470	2386	3878
WORD /dach/	12	80	806	1653	4616
VOICE /test/	59	167	390	2476	4462
WORD + VOICE dach/	8	167	736	1361	2897

• deviating (female) voice (VOICE) uttering the same word as that in the standard condition,

• simultaneous deviation in word and voice (WORD + VOICE)

The deviants were infrequent ($P = 0.15$; i.e., $P = 0.05$ for each deviant condition). Each deviant was preceded by at least three standards. The intensities were matched. The duration of both words was about 380 ms and the SOA was 900 ms. The average pitch was about 81 Hz for the male and 167 Hz for the female voice for both words. (For more detailed acoustic properties, see Table 1.)

Two sessions were held on two separate days. Each session contained four blocks of 20 min each. Altogether, 500 deviants and 9214 standards were presented. During the experiment, the subjects were instructed to watch a movie without sound and to ignore all auditory input.

Data Acquisition and Preprocessing

A 148-channel whole-head magnetometer system (MAGNES WHS-2500, 4D Neuroimaging) was used to record the MEG. The electrooculogram (EOG) was measured to identify epochs contaminated by eye artifacts. The head position was measured using five coils attached to the head which were localized by the system before and after each block. The MEG data were subjected to a band-pass filter of 1.5–20 Hz. This pass band is based on recommendations given by Sinkkonen and Tervaniemi (2000; see also Tervaniemi *et al.,* 1999), who stated that the MMN frequency range is between 1 and 20 Hz.

Since it is not possible to restore exactly a certain position of the dewar after the subject has moved, the data recorded from one subject during different sessions and blocks were first averaged within blocks (per condition) and then interpolated to a set of average sensor positions using a method based on linear inverse techniques (Knösche, 2002). As a result of this procedure, the values from the different blocks represented the magnetic field at the same positions with respect to the head and could be averaged first over blocks within each of the two sessions. The grand averages per session and condition are plotted in Fig. 1 to show the high degree of replicability in the data (sessions were on different days). Finally, the session av-

erages were combined to the subject averages, which were used for further analysis.

Data Analysis

The activation of the auditory cortices during this early processing of word and voice information was modeled using spatiotemporal dipole localization (Scherg and Berg, 1991). For this method, as for any source reconstruction scheme, we need a way to estimate measurements that will arise from an assumed source, the so-called forward solution. The forward solution in turn needs a model of the head that accounts for the different conductivities of the various head tissues. We employed the boundary element method (Ferguson *et al.,* 1994) consisting of one triangulated surface describing the inside of the skull (Hämäläinen and Sarvas, 1987). Such models can be generated from magnetic resonance images (MRI). In our case, a standardized head model, based on the Talairach scaled MRIs of 50 individuals, was employed. This standard model was then individually scaled to fit the shape of the subject's head as closely as possible in a least-squares sense. Six independent scaling factors ensured a close match of the resulting head shape. The head shape information was recorded together with three fiducial points (nasion, preauricular points) and the head position coils using a Polhemus FastTrak system.

The deviant stimuli in this study represent a discriminable change in unattended repetitive stimuli known to give rise to a mismatch negativity response in ERP experiments, usually peaking between 100 and 200 ms after stimulus onset (for an overview see Näätänen *et al.,* 2001). It has been found that the main generator of this component is located in the bilateral auditory cortex (Alho, 1995), where the exact location depends on the dimension of change (Giard *et al.,* 1995). Therefore, the analysis strategy and the employed dipole model concentrate on the characterization of the activity in both auditory cortices during the first 300 ms after stimulus onset.

First, we identified those subjects in whom the typical dipolar pattern of the N100m was clearly visible in both hemispheres, as described by Nakasato *et al.* (1995) (see also Fig. 1). Note that in contrast to the N100m elicited by simple beeps, the latency of the N100m following word stimulation is quite variable and in our case depends on the stimulus, hemisphere, and subject. For the standard condition the first negative component must be an N100m. Its latencies lie in the range of 130 to 160 ms (see also Fig. 4). For the deviant stimuli, the N100m is overlain by the mismatch field (MMF), which is the magnetic equivalent of the MMN. The latencies are in the expected range of 100 and 250 ms (see Näätänen, 2001). Eleven of the 19 subjects showed a clear N100m/MMF in all conditions. They were subjected to further analysis. The latencies

of the N100m/MMF component were compared statistically between the experimental conditions (stimuli) and brain hemispheres using variance analysis (ANOVA). For each subject and condition we defined a time window of 20 ms around the peak latency (averaged between the two hemispheres) and fitted a pair of fixed dipoles (Mosher and Leahy, 1998). Possible frontal contributions to the MMN as identified by Giard *et al.* (1990) using source current density maps were only represented by a current sink. Hence, their source orientation was likely to be radial and therefore could not contribute to the MEG. The resulting dipole positions were transformed to the standard brain using the scaling factors obtained previously from the individual scaling of the head models, averaged, and statistically compared using an ANOVA scheme.

In the next step, we investigated how the activity of the identified generators in the auditory cortices behaves during the first 300 ms after stimulus onset. For this purpose, position and orientation of the dipoles calculated in the previous step were kept fixed and only the magnitude was linearly fitted to each time sample of the MEG data. This strategy was justified, since the sources of the N100m, the MMF, and the next component, called P200m, are known to be just 1 cm apart (Rif *et al.,* 1991; Hari *et al.,* 1992) in the supratemporal plane and hence their field topologies are so similar that a projection of the later MEG activity onto the dipole localized from the N100m/MMF will yield a good qualitative estimation of the auditory cortices' activity.

Now, the resulting activation curves were realigned according to the latencies of the N100m/MMF components and then averaged over subjects. The averages of 10 time steps (spanning over 40 ms) centered on the N100m/MMF peak were tested in an ANOVA between conditions and hemispheres.

The realignment of the N100m/MMF peak also enabled us to assess the later activation independently; i.e., without the influence of the latency differences introduced by the N100m/MMF. The sources of this later activation were not localized because of the low signal-to-noise ratio in many subjects.

RESULTS

After localizing the dipoles and determining the time courses of their magnitudes, statistical tests were carried out to determine the dependency of dipole position, dipole strength, and peak latency on the experimental condition and the hemisphere. In Table 2 the results are summarized. For the dipole location, first an ANOVA of condition × hemisphere steps ($P < 0.05$, Huynh–Feldt correction) was computed for each of the three coordinates (*x*, left–right; *y,* posterior–anterior; *z,* inferior–superior). In the case of significant main effects, contrasts were computed between the different conditions. First, we tested our hypothesis that the

TABLE 2

Overview of the Effects Found on Dipole Location and Magnitudes as Well as the Latency of the N100m/MMF

	Dipole location			Dipole magnitude (N100m/MMF)	Peak latency (N100m/MMF)
	x	y	z		
2-factor ANOVA					
Main effect condition	—	*	—	**	*
Main effect hemisphere	*	**	—	**	—
Interaction condition × hemisphere	—	—	—	—	—
Contrasts					
STANDARD vs WORD	nt	*	nt	**	—
STANDARD vs VOICE	nt	**	nt	**	**
STANDARD vs WORD + VOICE	nt	**	nt	**	—
1-factor ANOVA between deviants	nt	—	nt	—	**
WORD + VOICE vs WORD	nt	nt	nt	nt	**
WORD + VOICE vs VOICE	nt	nt	nt	nt	**

Note. nt, not tested due to the lack of a main effect; —, $p > 0.05$.
  * $0.01 < p < 0.05$.
  ** $p < 0.01$.

deviants differ from the standard using t tests. We then computed a second ANOVA of condition (one factor of three steps) in order to investigate possible differences between the deviants. The same procedure was applied to the peak magnitudes of the dipole activation curves and the N100m/MMF peak latency. Since no interactions between hemisphere and condition were found, testing within each hemisphere separately was not legitimate.

Localization of N100m/MMF

The average locations of the generators of the N100m/MMF are depicted in Fig. 2. For the y axis (posterior–anterior), we found main effects of condition and hemisphere, but no interaction between the two factors. For the x axis (left–right), an effect of hemispheres was revealed. t tests between the conditions revealed that the source for the main effect on condition was the difference between each of the deviants and the standard. In fact, the deviants localized several millimeters more anteriorly than the standard dipole (see Fig. 2) and were not distinguishable from one another. Moreover, all dipoles in the right hemisphere were more anterior and lateral than those of the left hemisphere.

Activation Time Courses of Auditory Cortices

The activation time courses of the N100m/MMF dipoles (representing the generators in the auditory cortices) were computed after realignment of the time scales according to the latencies of the N100m/MMF peak. In Fig. 3, the results are summarized. For the statistical tests, see Table 2.

For the peak of the N100m/MMF, the ANOVA yielded significant main effects ($p < 0.01$) for both factors. However, no interaction between the factors could be found. In order to gain more insight into these effects, pair-wise t tests were carried out. As the sources of the main effect, the t tests clearly identified differences between all deviant conditions with respect to the standard condition, without any differences between the various deviants. Hence, we found that (1) the activation of the auditory cortices was stronger in the deviant conditions than in the standard conditions in both hemispheres and (2) the activation was stronger in the left than in the right hemisphere for all conditions.

Latencies of N100m/MMF

The average latencies of the identified N100m/MMF components are depicted in Fig. 4 together with their 95% confidence intervals. The ANOVA (Table 2) showed a significant main effect ($p < 0.05$) for condition, but no effect for hemisphere, nor any interaction between the factors. Further t tests between the deviants and the standard revealed that VOICE had a significantly longer latency than STANDARD. A single-factor ANOVA with a subsequent t test confirmed that the latency for the WORD + VOICE condition was significantly shorter than that for any of the other two deviants. Moreover, pair-wise f testing ($p < 0.005$, Bonferoni correction for multiple testing) revealed that the WORD + VOICE condition exhibited a significantly lower variance than that of all other conditions.

The last two rows of the diagram in Fig. 4 demonstrate that the latencies (average and confidence interval) found for the WORD + VOICE stimulus were very similar to the latencies predicted by a model of parallel processing of word and voice information (see Discussion).

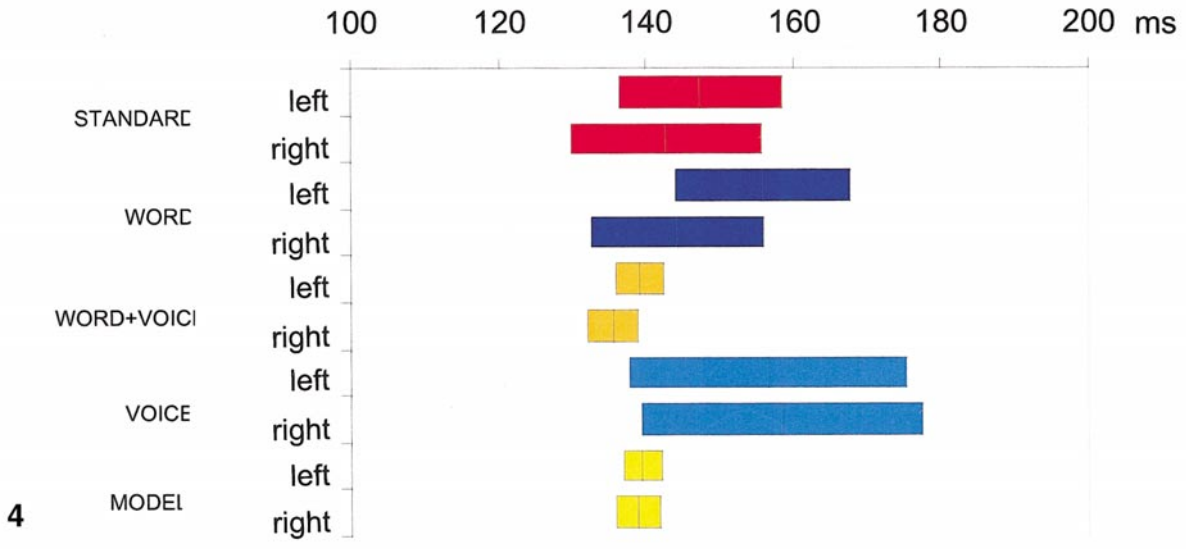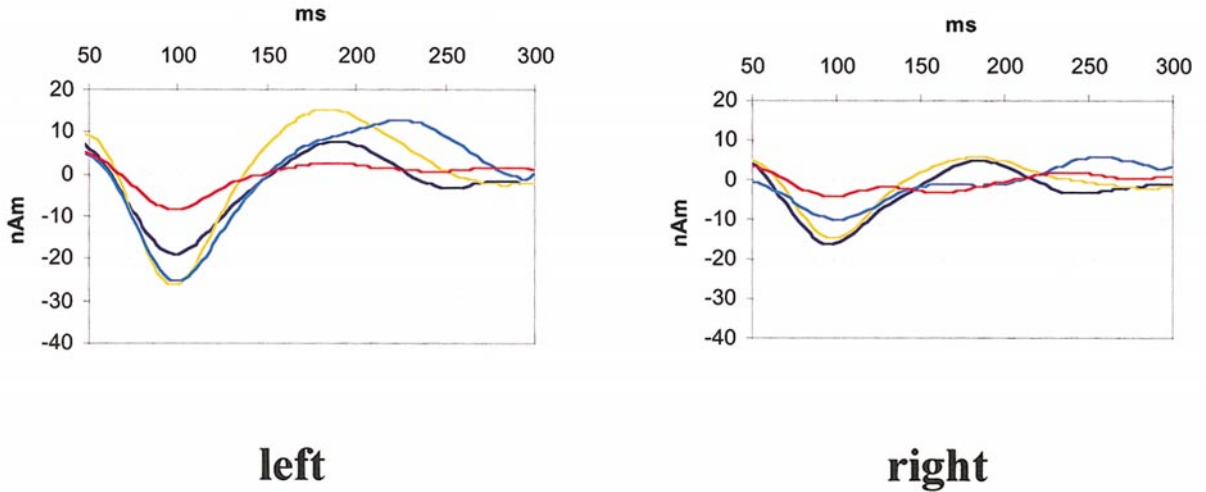Summarizing the most important findings, we can state that
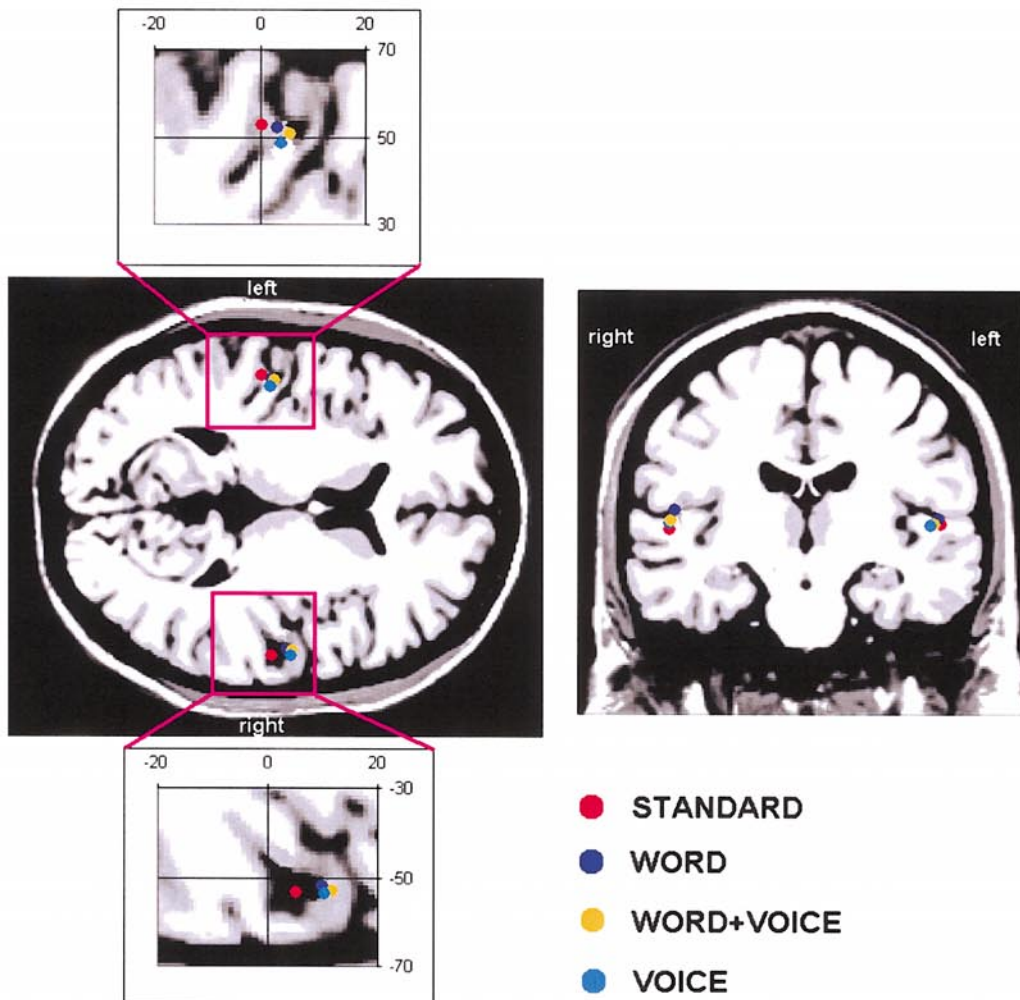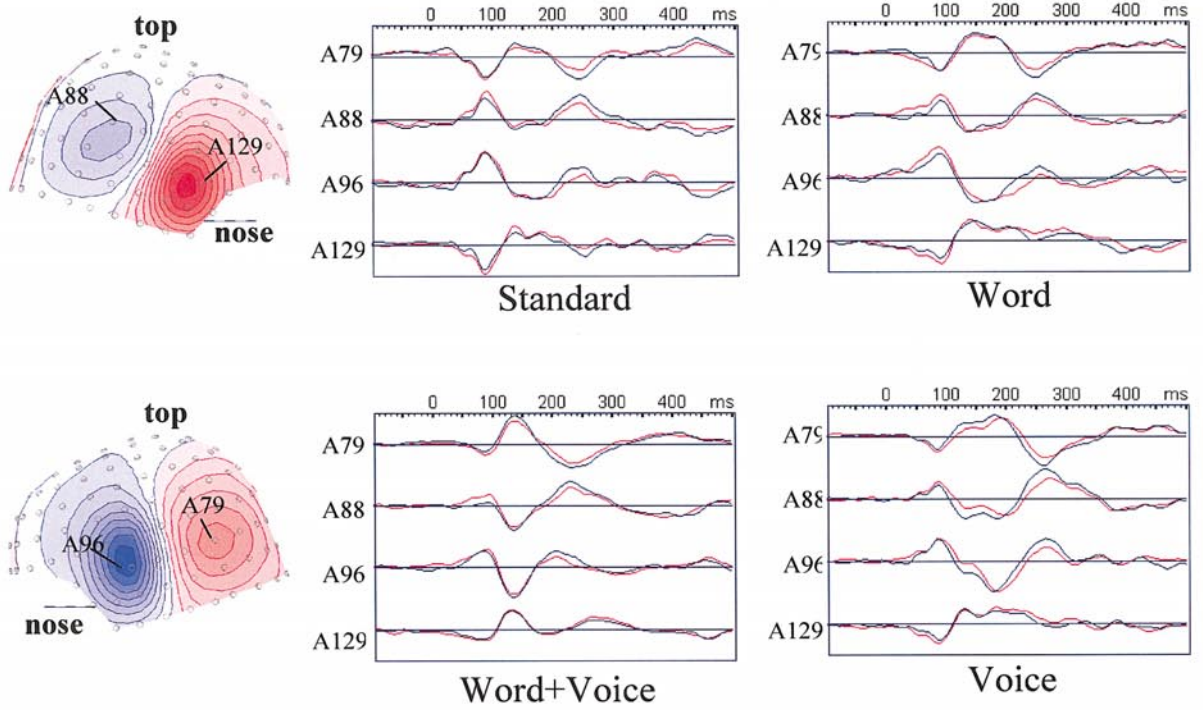
FIG. 3. Average activation time courses of the N100m/MMF dipole in both hemispheres after alignment of the individual curves according to the peak latency of the N100m/MMF.

FIG. 4. Average latencies of the N100m/MMF with 95% confidence intervals. The lowermost (yellow) bars represent the mathematical prediction of a integral parallel model for the conjunction of word and voice features.

(1) The processing of the deviants compared to the standard was deflected by a more anterior dipole position and by stronger dipole activation in both auditory cortices.

(2) The activity of the auditory cortices was stronger in the left than in the right hemisphere for all conditions.

(3) The location of the dipole representing auditory

FIG. 1. Identification of N100m by means of field patterns and replicability of recordings. Left column—field pattern seen from left and right, showing bilateral patterns of a downward-oriented dipole (blue, field pointing inward; red-field pointing outward). Middle and right column—traces for selected channels (approximate field maxima) and different sessions. The red curves represent the averages for the first session, the blue those for the second session.

FIG. 2. Average locations of the generators of the N100m, superimposed on the axial and coronal slices of the standard MRI, which pass through the center of mass of all depicted locations.
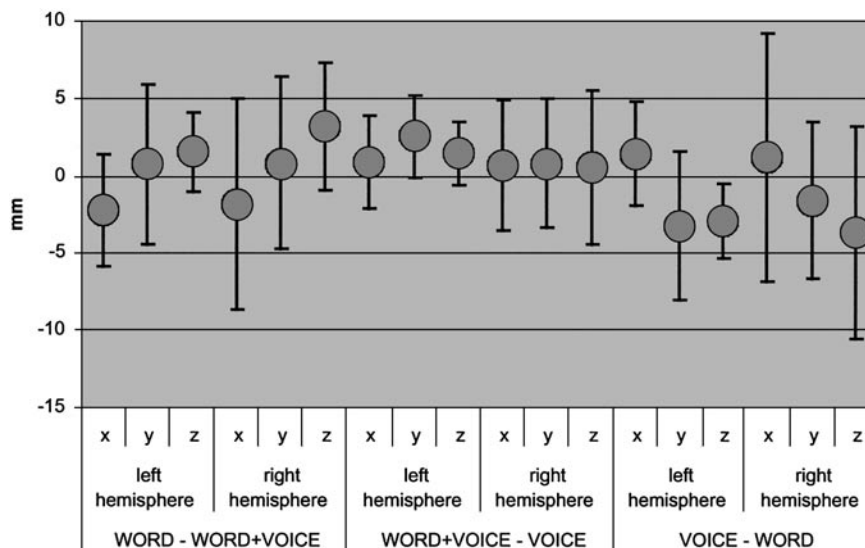
FIG. 5.   95% confidence intervals of the distances between the dipoles for the various deviant conditions.

cortex activity was more anterior and lateral in the right hemisphere than in the left one.

(4) Concerning the temporal structure of the N100m/MMF, there were differences between the conditions, with the combined deviant causing a significantly shorter and less variable latency compared to any other condition.

DISCUSSION

Hemispherical Specialization

This study was aimed at the spatial localization and hemispheric distribution of preattentive word and voice processing. In our study, the sources of the N100m/MMF following the deviant stimuli are different in location (more anterior) and magnitude (stronger) from the standard stimulus in both hemispheres. This is consistent with the simultaneous existence of the source of the N100m and a second, more anterior MMF source, as found in the literature (Csepe *et al.,* 1992). On the other hand, no differences in location or magnitude could be demonstrated between the various deviants. This is somewhat inconsistent with reports that the location of the MMF source depends on the dimension of change (Giard *et al.,* 1995; Alho *et al.,* 1996; FrodlBauch *et al.,* 1997; Diesch and Luce, 1997; Takegata *et al.,* 2001). However, most of these studies used nonspeech sounds, which were varied in pitch/ frequency, duration, intensity, or sound location. With respect to higher level parameters, Diesch and Luce (1997) presented evidence for the existence of different neuronal populations underlying the MMN elicited by different phonetic contrasts in consonant–vowel syllables. The dipole positions differed at most about 10 mm in the horizontal plane. Alho *et al.* (1998), however, did

not find any localization differences between different deviant syllables in an MMF experiment.

In our case, 95% confidence intervals of the position differences between the deviants (Fig. 5) show that statistically undetected differences of 3–10 mm are possible. Hence, our results cannot exclude the possibility of different neuronal populations being responsible for the detection of deviancy in different dimensions. They only prove that the distances between such areas must be small ($<$1 cm).

General hemispherical differences in dipole location were observed for all conditions. In particular, the dipoles in the right hemisphere were localized more anteriorly and laterally than in the left hemisphere. A similar result has been reported by Koyama *et al.* (2000) and, in particular in males, by Reite *et al.* (1995). This fact might point toward some general anatomical differences between the temporal lobes. Such differences have been found, especially pronounced in males, by volumetric analysis of MRIs (e.g., Penhune *et al.,* 1996; Good *et al.,* 2001). As the majority of the subjects in the present study were males, an anatomical cause for the hemispherical effect seems likely.

Finally, the activity strength of the auditory cortices, as represented by the dipole strength, exhibited a strong left lateralization for all conditions. There is a rich literature reporting hemispherical differences in electric or magnetic mismatch responses. Although the picture is not quite unanimous, it seems that the MMN/MMF elicited by simple tones is generally lateralized to the right hemisphere (Paavilainen *et al.,* 1991; Deouell *et al.,* 1998; Waberski *et al.,* 2001), while phonemes like consonant–vowel syllables give rise to a left-dominant mismatch response (Shtyrov *et al.,* 1998, 2000; Alho *et al.,* 1998, Rinne *et al.,* 1999). Since our

stimuli consisted of words of an even higher complexity than that of simple consonant–vowel syllables, the observed left laterality is in agreement with these data. Note that in our study not only the N100m/MMF but also the response to the standard stimuli was stronger in the left hemisphere, confirming the findings of Szymanski *et al.* (2001), who observed a left lateralization of middle latency auditory evoked fields in response to speech sounds, but not to simple tones.

Simultaneous Deviation in Several Dimensions

Another central question that this study tried to answer is whether word and voice information are processed sequentially or in parallel and, if processed in parallel, whether this is done *independently* or rather in a dependent, *integral* fashion. Since the latency of the N100m/MMF will play a dominant role in the following reasoning, we should first discuss whether any stimulus properties could have a direct influence on this parameter. In fact, it has been found by Sharma and Dorman (1993) that the latency of the N100 is influenced by the voice onset time (VOT) of the stimulus, but no such an effect is known for the MMN. Phillips *et al.* (2000) report an independence of the MMF from the acoustical properties of the stimulus. Since the component investigated here (N100m/MMF) is a superposition of the two, the question arises of whether the VOT has any decisive influence on the latencies. If this is the case, we should find systematic latency differences between WORD and WORD + VOICE on one side and VOICE and STANDARD on the other side (see Table 1). This, however, is not the case (see Table 2). Hence we conclude that any latency shifts introduced by the VOT must be small, at least in comparison to the effects discussed below.

Under the assumption of a *serial processing* of voice and word information, one would expect that either the two-dimensional deviation stimulus would lead to a longer processing time (if the mismatch response is seen only after both processes are finished) or the results would resemble the first of the two serial processes (if the mismatch response is seen after each of the deviancies has been processed).

In case of an *independent parallel* processing, the response to the combined stimulus should be the linear sum of the responses to the WORD and VOICE deviations, because it has been found in this study that the distance between the generators for the various N100m/MMF are quite small (<1 cm) and that their field topologies are therefore very similar. This would lead to an amplitude increase. Such additivity has already been shown for conjunctions of more basic stimulus features (interstimulus interval, duration, and frequency) by Levänen *et al.* (1993).

Instead, the latencies for the N100m/MMF were significantly shorter for the WORD + VOICE stimulus than for any of the other deviant stimuli. Moreover, the latency of the N100m/MMF exhibited much less inter-subject variance for the combined stimulus. Finally, the amplitudes of the responses to the combined stimuli could not be distinguished significantly from the amplitudes of the responses to simple deviations. Such effects can only be explained if one assumes *integral parallel* processes. For each of the dimensions of change, a separate unit extracts the features of the stimulus and compares them to the short-term memory traces. Differences will be integrated up and subjected to a threshold process triggering the deviancy detection, which is signaled by the mismatch response.

The exact mathematical prediction of this model is shown in the lowermost two rows of the diagram in Fig. 4. Clearly, the latencies as predicted by the model resemble the measured ones for the WORD + VOICE condition in terms of both expectation value and variance. Note that the latencies of the double-violation very much resemble the abovementioned effects of redundancy gain obtained in speeded-classification experiments (Mullenix and Pisoni, 1990; Wood, 1974). Thus, the observed MMFs may at least partially reflect the neural bases for the behavioral effects. Furthermore, it may be a general property of the MMN change detector that it is more efficient (less variance) and faster (shorter latency) to process a mismatch if it consists of combined deviations, as has been shown here for word and voice deviations.

CONCLUSIONS

The present study comprises the first neurophysiological investigation of the interaction of word and voice information. While behavioral studies (Wood, 1974; Mullenix and Pisoni, 1990) suggest a parallel processing of these two kinds of information at some point in the signal-evaluation process, the present data show that an interaction of the processes takes place even at preattentive perceptual stages within the auditory cortex.

For the processing of deviant stimuli, differing from a standard with respect to word information, voice information, or both we found a difference in source strength and source localization within the hemisphere (deviants were stronger and more anterior) as well as between the hemispheres (right was weaker and more anterior and lateral). The more anterior localization of the mismatch response with respect to the N100m/MMF for the deviants confirms findings in the literature. Lacking localization differences between the different types of deviants cannot be interpreted as proof for the sharing of the same neuronal substrate by all these processes, but demonstrates only that the distance between the centers of mass is smaller than 3–10 mm. The left dominance of the activation strength confirms findings in the literature that both N100m

and MMN caused by language stimuli are left dominant, in contrast to simple tones.

Interestingly, latency differences of the N100m/MMF were observed for the different deviants with a shorter and less variant latency for the combined stimulus, suggesting that word and voice information are processed in parallel. The proposed integral parallel model explains both latency shortening and a decrease of latency variance in the order of magnitude observed in the present study.

ACKNOWLEDGMENT

REFERENCES

Alho, K. 1995. Cerebral generators of mismatch negativity (MMN) and its magnetic counterpart (MMNm) elicited by sound changes. *Ear Hearing* **16:** 38–51.

Alho, K., Tervaniemi, M., Huotilainen, M., Lavikainen, J., Tiitinen, H., Ilmoniemi, R. J., Knuutila, J., and Näätänen, R. 1996. Processing of complex sounds in the human auditory cortex as revealed by magnetic brain responses. *Psychophysiology* **33:** 369–375.

Alho, K., Connolly, J. F., Cheour, M., Lehtokoski, A., Huotilainen, M., Virtanen, J., Aulanko, R., and Ilmoniemi, R. J. 1998. Hemispheric lateralization in preattentive processing of speech sounds. *Neurosci. Lett.* **258:** 9–12.

Belin, P., Zatorre, R. J., Lafalle, P., Ahad, P., and Pike, B. 2000. Voice-selective areas in human auditory cortex. *Nature* **403:** 309–312.

Binder, J. R., Frost, J. A., Hemmeke, T. A., Bellogwan, P. S. F., Springer, J. A., Kaufman, J. N., and Possing, E. T. 2000. Human temporal lobe activation by speech and nonspeech sounds. *Cereb. Cortex* **10:** 512–528.

Csepe, V., Pantev, C., Hoke, M., Hampson, S., and Ross, B. 1992. Evoked magnetic responses of the human auditory cortex to minor pitch changes—Localization of the mismatch field. *Electroencepalogr. Clin. Neurophysiol.* **84:** 538–548.

Deouell, L. Y., Bentin, S., and Giard, M. H. 1998. Mismatch negativity in dichotic listening: Evidence for interhemispheric differences and multiple generators. *Psychophysiology* **35:** 355–365.

Diesch, E., and Luce, T. 1997. Magnetic mismatch fields elicited by vowels and consonants. *Exp. Brain Res.* **116:** 139–152.

Elberling, C., Bak, C., Kofoed, B., Lebech, J., and Saermark, K. 1980. Magnetic auditory responses from the human brain—A preliminary report. *Scand. Audiol.* **9:** 185–190.

Ferguson, A. S., Zhang, X., and Stroink, X. 1994. A complete linear discretisation for calculating the magnetic field using the boundary element method. *IEEE Trans. Biomed. Eng.* **41:** 455–459.

FrodlBauch, T., Kathmann, N., Moller, H. J., and Hegerl, U. 1997. Dipole localization and test–retest reliability of frequency and duration mismatch negativity generator processes. *Brain Topogr.* **10:** 3–8.

Garner, W. R. 1974. *The Processing of Information and Structure.* Erlbaum, Hillsdale, NJ.

Giard, M. H., Perrin, F., Pernier, J., and Bouchet, P. 1990. Brain generators implicated in the processing of auditory stimulus deviance: A topographic event-related potential study. *Psychophysiology* **27:** 627–640.

Giard, M. H., Lavikainen, J., Reinikainen, K., Perrin, F., Bertrand, O., Pernier, J., and Näätänen, R. 1995. Separate representations

of stimulus frequency, intensity and duration in auditory sensory memory: An event-related potential and dipole-model analysis. *J. Cognit. Neurosci.* **7:** 133–143.

Good, C. D., Johnsrude, I., Asburner, J., Henson, R. N. A., Friston, K. J., and Frackowiak, R. S. 2001. Cerebral symmetry and the effects of sex and handedness on brain structure: A voxel-based morphometric analysis of 465 normal adult human brains. *NeuroImage* **14:** 685–700.

Hämäläinen, M. S., and Sarvas, J. 1987. Feasibility of the homogeneous head model in the interpretation of neuromagnetic data. *Phys. Med. Biol.* **32:** 91–97.

Hari, R., Attoniemi, K., Jarvinen, M. L., Katila, T., and Varpula, T. 1980. Aditory evoked transient and sustained magnetic fields of the human brain—Localization of neural generators. *Exp. Brain Res.* **40:** 237–240.

Hari, R., Hämäläinen, M., Ilmoniemi, R., Kaukoranta, E., Reinikainen, K., Salminin, J., Alho, K., Näätänen, R., and Sams, M. 1984. Response of the primary auditory cortex to pitch changes in a sequence of tone pips: Neuromagnetic recordings in man. *Neurosci. Lett.* **50:** 127–132.

Hari, R., Rif, J., Tiihonen, J., and Sams, M. 1992. Neuromagnetic mismatch fields to single and paired tones. *Electroencephalogr. Clin. Neurophysiol.* **82:** 152–154.

Imaizumi, S., Mori, K., Kiritani, S., Hosoi, H., and Tonoike, M. 1998. Task-dependent laterality for cue decoding during spoken language processing. *Neuroreport* **9:** 899–903.

Kasai, K., Yamada, H., Kamio, S., Nakagome, K., Iwanami, A., Fukuda, M., Itoh, K., Koshida, I., Yumoto, M., Iramina, K., Kato, N., and Ueno, S. 2001. Brain lateralization for mismatch response to across- and within-category change of vowels. *Neuroreport* **12:** 2467–2471.

Knösche, T. R. 2002. Transformation of whole-head MEG recordings between different sensor positions. *Biomed. Eng.* **47:** 59–62.

Koyama, S., Gunji, A., Yabe, H., Oiwa, S., Akahane-Yamada, R., Kakigi, R., and Näätänen, R. 2000. Hemispheric lateralization in an analysis of speech sounds—Left hemisphere dominance replicated in Japanese subjects. *Cognit. Brain Res.* **10:** 119–124.

Levänen, S., Hari, R., McEvoy, L., and Sams, M. 1993. Responses of the human auditory cortex to changes in one vs two stimulus features. *Exp. Brain Res.* **97:** 177–183.

Levänen, S., Ahonen, A., Hari, R., McEvoy, L., and Sams, M. 1996. Deviant auditory stimuli activate human left and right auditory cortex differently. *Cereb. Cortex* **6:** 288–296.

Levy, D. A., Granot, R., and Bentin, S. 2001. Processing specificity for human voice stimuli: Electrophysiological evidence. *Neuroreport* **12:** 2653–2657.

Liasis, A., Towell, A., and Boyd, S. 2000. Intracranial evidence for differential encoding of frequency and duration discrimination responses. *Ear Hearing* **21:** 252–256.

Mathiak, K., Hertrich, I., Lutzenberger, W., and Ackermann, H. 2000. Encoding of temporal speech features (formant transients) during binaural and dichotic stimulus application: A whole-head magnetoencephalography study. *Cogn. Brain Res.* **10:** 125–131.

Mosher, J. C., and Leahy, R. M. 1998. Recursive music—A framework for EEG and MEG source localization. *IEEE Trans. Biomed. Eng.* **45:** 1342–1354.

Mullenix, J. W., and Pisoni, D. B. 1990. Stimulus variability and processing dependencies in speech perception. *Percep. Psychophys.* **47:** 379–390.

Näätänen, R., Gaillard, A. W. K., and Mantysalo, S. 1978. Early selective-attention effect on evoked potential reinterpreted. *Acta Psychol.* **42:** 313–329.

Näätänen, R. 1992. *Attention and Brain Function.* Erlbaum, Hillsdale, NJ.

Näätänen, R., Paavilainen, P., Tiitinen, H., Jiang, D., and Alho, K. 1993. Attention and Mismatch Negativity. *Psychophysilogy* **30:** 436–450.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, R. J., Luuk, A., Allik, J., Sinkkonen, J., and Alho, K. 1997. Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature* **385:** 432–434.

Näätänen, R. 2001. The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology* **38:** 1–21.

Näätänen, R., Tervaniemi, M., Sussman, E., Paavilainen, P., and Winkler, I. 2001. "Primitive intelligence" in the auditory cortex. *Trends Neurosci.* **24:** 283–288.

Nakasato, N., Fujita, S., Seki, S., Kawamura, T., Matani, A., Tamura, I., Fujiwara, S., and Yoshimoto, T. 1995. Functional localization of bilateral auditory cortices using an MRI-linked whole-head magnetoencephalography (MEG) system. *Electroencephogr. Clin. Neurophysiol.* **94:** 183–190.

Paavilainen, P., Alho, K., Reinikainen, K., Sams, M., and Näätänen, R. 1991. Right-hemisphere dominance of mismatch negativities. *Electroencephogr. Clin. Neurophysiol.* **78:** 466–479.

Paavilainen, P., Valppu, S., and Näätänen, N. 2001. The additivity of the auditory feature analysis in the human brain as indexed by the mismatch negativity: $1 + 1 \approx 2$ but $1 + 1 + 1 < 3$. *Neurosci. Lett.* **301:** 179–182.

Pantev, C., Hoke, M., Lehnertz, K., Lütkenhöner, B., Anogianakis, G., and Wittowski, W. 1988. Tonotopic organization of the human auditory cortex revealed by transient auditory evoked magnetic fields. *Electroencephogr. Clin. Neurophysiol.* **69:** 160–170.

Penhune, V. B., Zatorre, R. J., MacDonald, J. D., and Evans, A. C. 1996. Interhemispheric anatomical differences in human primary auditory cortex: Probabilistic mapping and volume measurement from magnetic resonance scans. *Cereb. Cortex* **6:** 661–672.

Phillips, M. D., Pellathy, T., Marantz, A., Yellin, E., Wexler, K., Poeppel, D., McGinnis, M., and Roberts, T. 2000. Auditory cortex accesses phonological categories: An MEG mismatch study. *J. Cognit. Neurosci.* **12:** 1038–1055.

Reite, M., Sheeder, J., Teale, P., Richardson, D., Adams, M., and Simon, J. 1995. MEG based brain laterality: Sex differences in normal adults. *Neuropsychologia* **33:** 1607–1616.

Rif, J., Hari, R., Hämäläinen, M. S., and Sams, M. 1991. Auditory attention affects 2 different areas in the human supratemporal cortex. *Electroencephogr. Clin. Neurophysiol.* **79:** 464–472.

Rinne, T., Alho, K., Alku, P., Holi, M., Sinkkonen, J., Virtanen, J., Bertrand, O., and Näätänen, R. 1999. Analysis of speech sounds is left-hemisphere pre-dominant 150 ms after sound onset. *Neuroreport* **10:** 1113–1117.

Sams, M., Paavilainen, P., Alho, K., and Näätänen, R. 1985. Auditory frequency discrimination and event-related potentials. *Electroencephogr. Clin. Neurophysiol.* **62:** 437–448.

Scherg, M., and Berg, P. 1991. Use of prior knowledge in brain electromagnetic source analysis. *Brain Topography* **4:** 143–150.

Sharma, A., and Dorman, M. F. 1993. Cortical auditory evoked potential correlates of categorical perception of VOT. *J. Acoust. Soc. Am.* **106:** 1078–1083.

Shtyrov, Y., Kujala, T., Ahveninen, J., Tervaniemi, M., Alku, P., Ilmoniemi, R. J., and Näätänen, R. 1998. Background acoustic noise and the hemispheric lateralization of speech processing in the human brain: Magnetic mismatch negativity study. *Neurosci. Lett.* **251:** 141–144.

Shtyrov, Y., Kujala, T., Lyytinen, H., Kujala, J., Ilmoniemi, R. J., and Näätänen, R. 2000. Lateralization of speech processing in brain as indicated by mismatch negativity and dichotic listening. *Brain Cognit.* **43:** 392–398.

Sinkkonen, J., and Tervaniemi, M. 2000. Towards optimal recording and analysis of the mismatch negativity. *Audiol. Neuro-Otol.* **5:** 235–246.

Szymanski, M. D., Perry, D. W., Gage, N. M., Rowley, H. A., Walker, J., Berger, M. S., and Roberts, T. P. L. 2001. Magnetic source imaging of late evoked field responses to vowels: Toward an assessment of hemispheric dominance for language. *J. Neurosurg.* **94:** 445–453.

Takegata, R., Syssoeva, O., Winkler, I., Paavilainen, P., and Näätänen, R. 2001. Common neural mechanism for processing onset-to-onset intervals and silent gaps in sound sequences. *Neuroreport* **12:** 1783–1787.

Tervaniemi, M., Lehtokoski, A., Sinkkonen, J., Virtanen, J., Ilmoniemi, R. J., and Näätänen, R. 1999. Test–retest reliability of mismatch negativity for duration, frequency and intensity changes. *Clin. Neurophysiol.* **110:** 1388–1393.

Titova, N., and Näätänen, R. 2001. Preattentive voice discrimination by the human brain as indexed by the mismatch negativity. *Neurosci. Lett.* **308:** 63–65.

Van Lancker, D., and Kreiman, J. 1987. Voice discrimination and recognition are separate abilities. *Neuropsychologia* **25:** 829–834.

Van Lancker, D., Cummings, J. L., Kreiman, J., and Dobkin, B. H. 1988. Phonagnosia: A dissociation between familiar and unfamiliar voices. *Cortex* **24:** 195–209.

Van Lancker, D., Kreiman, J., and Cummings, J. 1989. Voice perception deficits: Neuroanatomical correlates of phonagnosia. *J. Clin. Exp. Neuropsychol.* **11:** 665–674.

Waberski, T. D., Kreitschmann-Andermahr, I., Kawohl, W., Darvas, F., Ryang, Y., Gobbele, R., and Buchner, H. 2001. Spatio-temporal source imaging reveals subcomponents of the human auditory mismatch negativity in the cingulum and right inferior temporal gyrus. *Neurosci. Lett.* **308:** 107–110.

Winkler, I., Paavilainen, P., and Näätänen, R. 1992. Can echotic memory store 2 traces simultaneously–A study of event-related brain potentials. *Psychophysiol.* **29:** 337–349.

Wolff, C., and Schröger, E. 2001. Human pre-attentive auditory change-detection with single, double, and triple deviations as revealed by mismatch negativity additivity. *Neurosci. Lett.* **311:** 37–40.

Wood, C. 1974. Parallel processing of auditory and phonetic information in speech discrimination. *Percept. Psychophys.* **17:** 501–508.

Yvert, B., Crouzeix, A., Bertrand, O., Seither-Preisler, A., and Pantev, C. 2001. Multiple supratemporal sources of magnetic and electric auditory evoked middle latency components in humans. *Cereb. Cortex* **11:** 411–423.