◆ Human Brain Mapping 20:13–21(2003) ◆

Dissociation of Human and Computer Voices in the Brain: Evidence for a Preattentive Gestalt-Like Perception

Sonja Lattner, Burkhard Maess,* Yunhua Wang, Michael Schauer, Kai Alter, and Angela D. Friederici

Max-Planck-Institute of Cognitive Neuroscience, Leipzig, Germany

Abstract: We investigated the early ("preattentive") cortical processing of voice information, using the so-called "mismatch response." This brain potential allows inferences to be made about the sensory short-term store. Most importantly, the mismatch potential also provides information about the organization of long-term memory traces in the auditory system. Such traces have reliably been reported for phonemes. However, it is unclear whether they also exist for human voice information. To explore this issue, 10 healthy subjects were presented with a single word stimulus uttered by voices of different prototypicality (natural, manipulated, synthetic) in a mismatch experiment (stimulus duration 380 msec, onset-to-onset interval 900 msec). The event-related magnetic fields were recorded by a 148-channel whole-head magnetometer and a source current density modeling of the magnetic field data was performed using a minimum-norm estimate. Each deviating voice signal in a series of standard-voice stimuli evoked a mismatch response that was localized in temporal brain regions bilaterally. Increased mismatch related magnetic flux was observed in response to decreased prototypicality of a presented voice signal, but did not correspond to the acoustic similarity of standard voice and deviant voices. We, therefore, conclude that the mismatch activation predominantly reflects the ecological validity of the voice signals. We further demonstrate that the findings cannot be explained by mere acoustic feature processing, but rather point towards a holistic mapping of the incoming voice signal onto long-term representations in the auditory memory. *Hum. Brain Mapping 20:13–21, 2003.* © 2003 Wiley-Liss, Inc.

Key words: auditory evoked fields; cortical representations; long-term memory; mismatch negativity; speech perception; magnetoencephalography; MEG

INTRODUCTION

The mismatch negativity (MMN) event-related potential reflects the preattentive cortical processing of a change in some constant aspect of the auditory input [Näätänen et al., 1978; Näätänen et al., 2001; Näätänen, 1992]. Using an oddball procedure, i.e., frequent standard stimuli and infrequent deviant stimuli, the amplitude of the MMN is known to increase with decreasing physical similarity between deviant and standard. However, while this relationship between physical similarity and MMN amplitude has proven valid for sinusoidal tones [Hari et al., 1980; Näätänen et al., 1978; Sams et al., 1985] and complex tones [Näätänen et al., 1993; Vihla et al., 2000], it has been

Contract grant sponsor: Leibniz-Preis.

*Correspondence to: Burkhard Maess, Stephanstr. 1a, D-04103 Leipzig, Germany. E-mail: maess@cns.mpg.de

Received for publication 11 October 2002; accepted 24 April 2003
DOI 10.1002/hbm.10118

shown that the mismatch negativity in response to deviating speech sounds is dependent on language-specific long-term memory traces. Non-prototypical speech sounds in a given language, i.e., foreign phonemes, evoke an MMN that is weaker than the one evoked by the prototypical phonemes [Näätänen et al., 1997; Winkler et al., 1999]. These findings indicate that the MMN not only reflects the physical properties of a given input, but also the higher cognitive properties of the speech processing system. Therefore, the MMN has been described as "primitive intelligence in the auditory cortex" [Näätänen et al., 2001].

However, during speech perception, listeners not only decode speech sounds in order to understand the speaker's intended message, but in addition, they consistently extract information as to the speaker's age, gender, and other characteristics from the voice signal. We will label this type of information "voice information". In terms of acoustic properties, it has been described by two main parameters: (1) the fundamental frequency, and (2) the timbre, i.e., the characteristic spectral frequencies or formants for each speaker [Fant, 1960]. The auditory processing of voice information is one of the most important human abilities, in phylogenesis as well as in everyday communication [Mann et al., 1979]. Only recently have functional studies investigated the neural basis of voice processing. Using functional magnetic resonance imaging (fMRI) it was shown that the perception of human vocalizations can be dissociated from the perception of other sounds by a characteristic neural response in so-called "voice-selective brain areas" that are distributed along the superior temporal sulcus STS [Belin et al., 2000, 2002]. One focus of activation has been localized near the core areas of the auditory cortex (A1), which might suggest a sensitivity for voice information already at early processing stages [Belin et al., 2002]. A recent electrophysiological experiment has demonstrated that a change of speaker is detected within a few hundred milliseconds. An increasing physical difference between a standard and deviant human voices led to an increase in the amplitude of the mismatch negativity [Titova and Näätänen, 2001], comparable to what has been found in tone processing. However, so far, only prototypical voices have been investigated. Therefore, it is unclear whether the reported mismatch effect reflects unspecific auditory processing in short-term sensory memory or whether it reflects early voice-specific processing, i.e., whether it is dependent on long-term memory traces of voice prototypes, comparable to what has been found for speech sound perception. The aim of the present study was, therefore, to examine the preattentive perception of prototypical (human) and non-prototypical (computer manipulated) voices in order to find out whether voice information is processed in voice-specific ways.

SUBJECTS AND METHODS

Stimuli and stimulus rating

The experiment consisted of four conditions and each condition corresponded to one experimental stimulus. All stimuli consisted of utterances of the high-frequent concrete German noun *Dach* (roof). Another study comparing the effect of voice vs. word deviancy was published recently [Knösche et al., 2002]. In one condition, labeled MALE, the stimulus was uttered by a male voice. In another condition, labeled FEM, it was uttered by a female speaker. These two stimuli will henceforth be termed prototypical voices because they comprise unmanipulated human speech. In contrast, the remaining two speech signals consisted of non-prototypical voices: The stimulus in condition PITCH was the same stimulus as used in condition FEM, but its fundamental frequency (F0) was shifted until it matched the pitch of the male voice maintaining the original F0 contour. This F0 shift was performed using the PSOLA resynthesis tool of the speech editor PRAAT (http://fonsg3.uva.nl/praat). The fourth condition, labeled SYN, comprised a stimulus that was synthesized using the MBROLA diphone synthesizer (http://tcts.fpms.ac.be/synthesis/; German database female voice) on the model of stimulus in condition FEM. Acoustic analyses showed that this synthesized stimulus comprised a number of spectral peculiarities, leading to the impression of a "computer voice". Most striking was a lack of energy in the 250-Hz range as well as an additional energy band in higher frequencies (see Fig. 1; the cochleagram was modeled by the PRAAT speech editor at a window size of 0.003 sec; and the spectrum was calculated at a window size of 0.002 sec). Further acoustic properties of all stimuli are listed in Table I. In the study, the stimulus MALE was employed as frequent standard stimulus, whereas all other stimuli were used as deviants. It is, therefore, important to note that all female voices (conditions FEM, PITCH, SYN) differed from the male voice in timbre, i.e., in the formant structure, but the stimuli FEM and SYN also differed in pitch-information, whereas the stimulus PITCH did not. Therefore, the stimulus PITCH was acoustically more similar to the stimulus MALE than the other stimuli. The stimuli were sampled at 16 kHz, the intensity of the stimuli was normalized using the am-

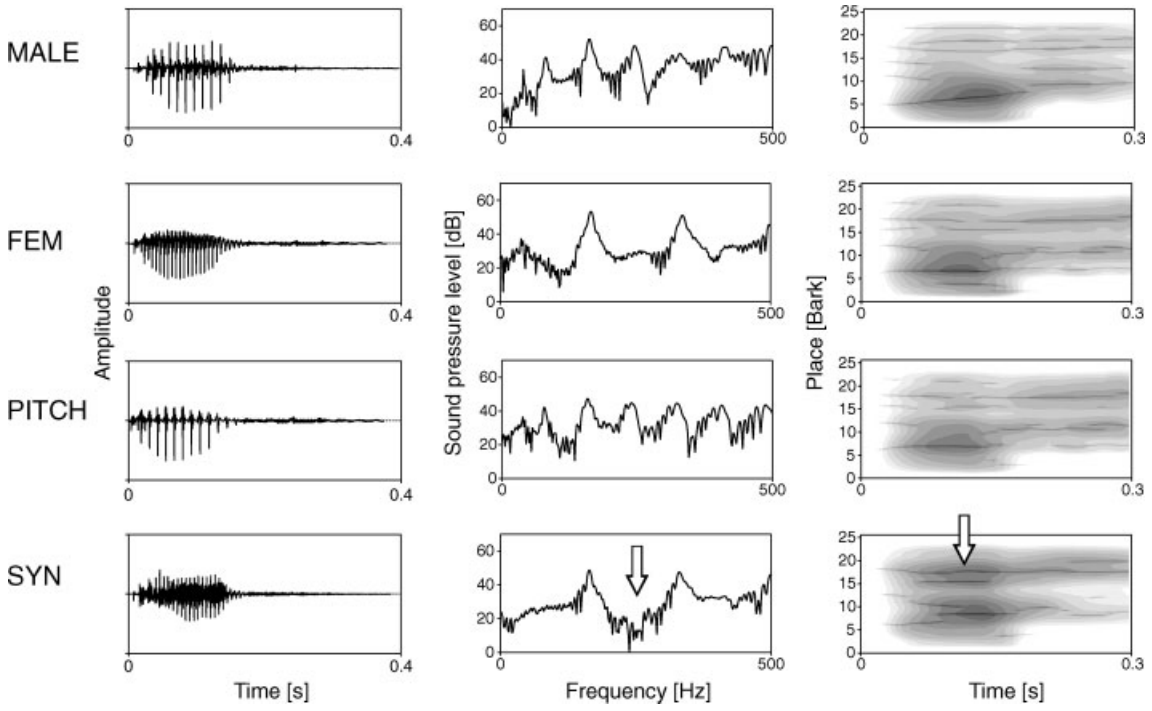

Figure 1.

Physical properties of the stimuli: waveform display (left), spectrum (middle) and cochleagram (right) of the male voice (MALE), the female voice (FEM), the pitch-shifted stimulus (PITCH) and the synthesized speech signal (SYN). The stimuli were of the same duration. The stimuli MALE and PITCH have a low fundamental frequency (FO). Arrows indicate the unusual spectral properties of the artificial signal, a lack in the 300 Hz range and an additional energy band in the overtone regions of about 16–18 Bark.

plitude-based normalizing procedure of the Cool Edit software (Syntrillium Software Corporation, Phoenix, AZ). For ease of presentation, we will refer to the stimuli by the name of the given condition, i.e., MALE, FEM, PITCH, and SYN.

Rating

To better understand the perceptual relevance of the independent variable (prototypicality of the stimuli), a rating study was performed. Two groups of linguistically naive listeners who did not participate in the later MEG experiments took part. A first group of subjects (n = 15, six male, age 19–39 years) rated the stimuli with respect to naturalness/pleasantness. At a familiarization stage, the subjects were presented with the stimuli twice (inter-stimulus-interval, ISI 3,000 msec). On the third presentation, they had to rate the "naturalness" of the voices on a five-point scale (1 = natural; 5 = unnatural). As the term "naturalness" is rather vaguely defined and not universally accepted [Nusbaum et al., 1995; 2000], listeners were asked to additionally rate the pleasantness of the voices on a five-point scale (1 = very pleasant; 5 = very unpleasant).

TABLE I. Physical properties of the stimuli: mean frequencies of the first four formants*

Stimulus	VOICE	F0	F1	F2	F3
MALE	Male	80	806	1,653	4,616
FEM	Female	167	736	1,361	2,897
PITCH	F0-shifted	80	718	1,328	2,777
SYN	Synthesized	167	856	1,308	2,824

* Values represent Hz.

The "naturalness" rating revealed a difference between the speech signals of human origin (MALE, FEM), which were judged as very natural (<2 points) and those that were manipulated (PITCH) or synthesized (SYN); *t* tests showed that the difference between these clusters was significant (t_{14}=9.37, $P < 0.0001$). There was neither a significant difference between the stimuli MALE and FEM ($t_{14} = 1.29$, $P < 0.22$) nor between PITCH and SYN ($t_{14} < 1$). In the "pleasantness" rating, the stimuli MALE and FEM were both rated as highly pleasant (<2 points; no

significant difference). The pitch-manipulated stimulus (PITCH) gained a middle score of 3.9 points, differing significantly from the natural voices (t_{14}=9.32, $P < 0.0001$). The synthesized stimulus SYN was rated as most unpleasant (4.5 points), differing significantly from PITCH ($t_{14} = 2.26$, $P < 0.05$).

The second group of listeners (n=12, five male, age 20–30 years) rated the "similarity" of the voices on a five-point scale (1 = very similar; 5 = very dissimilar). Here, we report the similarity of any of the female voices (FEM, PITCH, SYN) that were used as deviants in the following MEG study, and the male voice (MALE) that was used as the standard stimulus. The natural female voice was perceived as being only marginally less similar to the male voice (3.6 points) than the pitch-shifted stimulus (PITCH, 3.2 points). The stimulus SYN was rated as most different from stimulus MALE (4.5 points). However, only the difference of PITCH vs. SYN reached statistical significance in the similarity rating ($t_{14} = 3.36$, $P < 0.01$). Subjects found this rating task very hard to perform. Post-hoc, we attribute these difficulties to the conflict of acoustic similarity and voice prototypicality.

MEG experiment

Predictions

Employing the stimuli in a standard oddball paradigm in combination with an MEG measurement allowed us to test two mutually exclusive hypotheses. A first hypothesis holds that the mismatch response to voice deviants reflects non-specific auditory processing. In this case, the MMNm should mirror the physical deviation between the standard stimulus and the three female voice deviants: the stimulus PITCH should evoke the weakest response because it is comprised of the same fundamental frequency as the standard stimulus. A medium response would be expected for the natural female voice FEM, whereas the artificial speech signal SYN, deviating most strongly from the standard, should evoke the strongest MMNm. A second hypothesis is based on the assumption that the mismatch response is dependent on long-term auditory memory traces of average voices. In this case, the MMNm to the deviants should exhibit a pattern that relates to the naturalness or pleasantness rating with stimulus FEM evoking a weak response and the two unnatural voices SYN and PITCH evoking significantly stronger mismatch responses.

Subjects

Ten right-handed, native German speaking adults (age 21–30 years; 4 males) participated in the MEG experiment. Subjects belonged to the subject pool of the Max-Planck Institute of Cognitive Neuroscience Leipzig and gave written informed consent in accordance with the guidelines approved by the Ethics Committee of the Leipzig University Medical Faculty.

Procedure

The subject was seated in a comfortable chair beneath the dewar of the wholehead magnetometer. At approximately 1.20-m distance a silent movie was presented on a projection screen. Stimuli were binaurally presented via eartubes. An oddball design was employed whereby the male voice (MALE) served as frequent standard stimulus and the three female voices (stimuli FEM, PITCH, SYN) as infrequent deviants. Five hundred deviants of each type and 9,200 standards were presented in a pseudo-randomized order, allowing a deviant after at least three standards. In the statistic analysis, however, only standard epochs immediately preceding a deviant were considered. The onset-to-onset interval was 900 msec. Because of the relatively long duration of the experiment, the data were acquired at two sessions of 90 min each, held on two separate days.

MEG data processing

The auditory evoked magnetic fields were recorded by a 148-channel whole-head magnetometer (MAGNES WHS 2500; 4D Neuroimaging, San Diego, CA). The data were sampled at about 254 Hz (recording bandpass 0.1 to 50 Hz) and off-line filtered from 1.5–20.0 Hz. This band-pass filter was based on recommendations of Sinkkonen and Tervaniemi, stating that the MMN frequency range lies between 1–20 Hz. The MEG signal was epoched for a 700-msec period (−200 to 500 msec with respect to the stimulus onset; the −100 to 0 time window served as a baseline). EOG was measured and epochs containing eyeblinks or other artifacts such as channel drifts were rejected. Subjects' head positions as determined by anatomical reference points were recorded per session. Since it is not possible to restore the positions of dewar and head once the subject has moved, the data of each subject were first averaged for each block and prior to further processing normalized in relation to a standard head orientation. For each subject, a source current density (SCD) modeling was performed using the minimum-

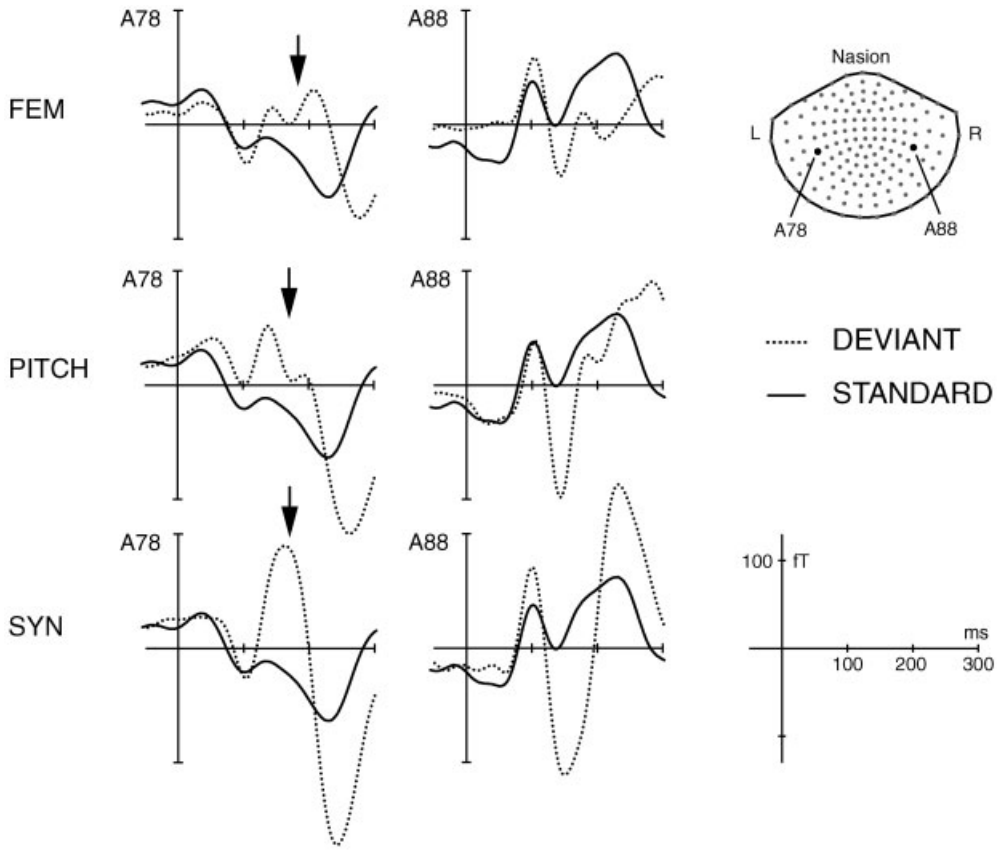

Figure 2.

Magnetic field strengths from two selected channels near the auditory cortices of the left (channel 78) and right (channel 88) hemisphere. Compared to the standard stimulus (solid line) deviants (dotted line) evoked a mismatch negativity indicated by the arrows; note the typical inversion of the magnetic field direction in the other hemisphere.

norm least-square algorithm [Hämälainen and Ilmoniemi, 1994; Wang et al., 1992] of the CURRY 4.0 software (Phillips, Germany). A Boundary Element Model (BEM) based on the anatomical scans of 50 healthy adults [Ferguson et al., 1994] was used as a volume conductor. Responses between 100 and 300 msec post-onset of a deviant were accepted as mismatch-related activity and averaged for each condition and hemisphere. Note that we did not perform a subtraction of deviant and standard conditions, because they were associated with physically different stimuli. Therefore, we prefer to label the observed magnetic field mismatch response rather than MMNm.

RESULTS

The signals from two channels in the center of the evoked magnetic fields (near the left and right auditory cortex) were selected. The auditory responses for each deviant were compared with the response to the standard stimulus (condition MALE). For each deviant, a mismatch response in the magnetic fields was observed. As displayed in Figure 2, the mismatch effect for the two selected channels was maximal in the 100–200-msec time range. Therefore, paired *t*-tests of the magnetic field strength from 100–200 msec were performed for each deviant condition against the standard condition MALE. For each deviant, there was a mismatch response in the magnetic field pattern that differed significantly from the responses to the standard stimulus. (Channel 78: SYN $t_9 = 9.7$, $P < 0.0001$; PITCH $t_9 = 10.76$, $P < 0.0001$; FEM $t_9 = 4.09$, $P < 0.01$); Channel 88: SYN $t_9 = 5.73$, $P < 0.001$; PITCH $t_9 = 3.96$, $P < 0.01$; FEM $t_9 = 2.23$, $P = 0.052$).

However, of main interest was the comparison of the source current density (SCD) in response to the various deviants (Fig. 3).

Most strikingly, there was an increase in the activity that corresponded inversely to the signals prototypicality. The average current density values taken from 100–

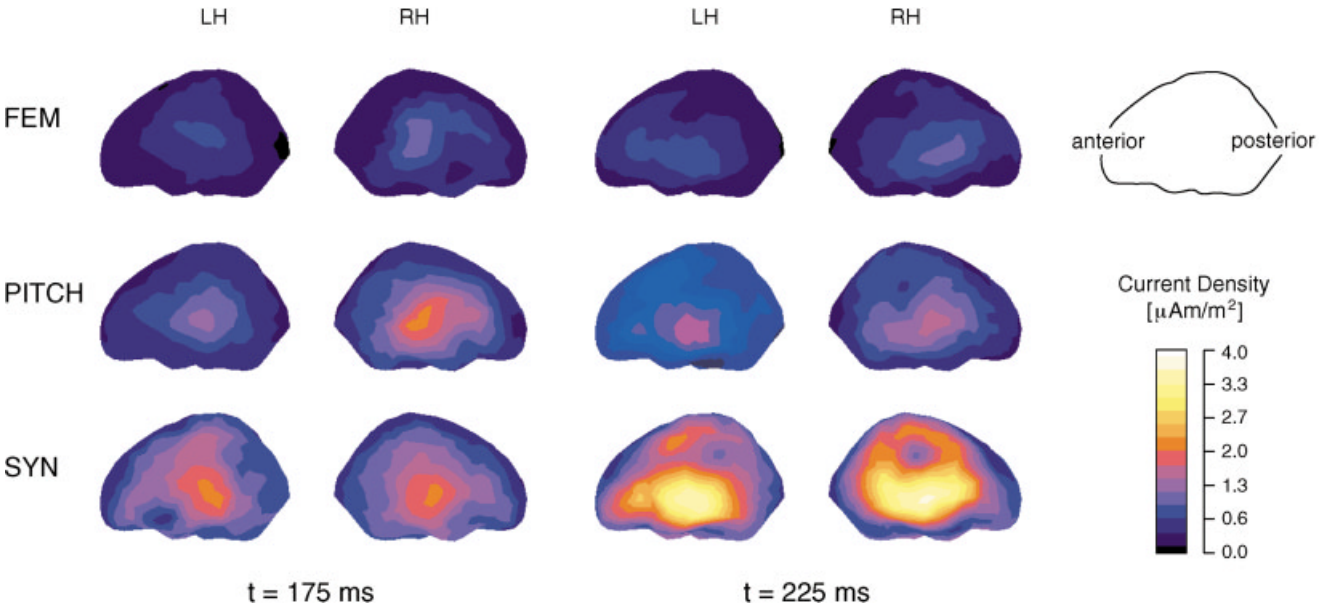

Figure 3.

Source current density maps of left hemisphere (LH) and right hemisphere (RH) at the time points t = 175 ms and t = 225 ms after stimulus onset, in response to the three deviants FEM (natural female voice), PITCH (pitch-shifted voice) and SYN (synthesized voice).

300 msec post stimulus-onset were subject to a two-way ANOVA of the factors Deviant (FEM, PITCH, SYN), and Hemisphere (left, right). The analysis showed a significant main effect of Deviant $F_{2,18} = 19.21$, $P < 0.0001$; see Fig. 4). There was no main effect of Hemisphere and no interaction between the two factors. The main effect of Deviant was resolved according to the stimulus prototypicality. The two *t*-tests (conditions FEM vs. PITCH; conditions PITCH vs. SYN) revealed a significant difference between the conditions FEM and PITCH (t_9=4.15, $P < 0.01$) as well as between the conditions PITCH and SYN (t_9 =3.06, $P < 0.05$).

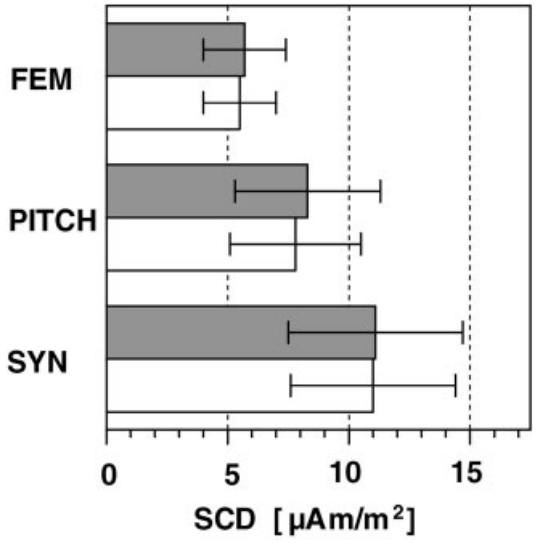

Figure 4.

Average source current density strength from 100–300 ms after stimulus onset for each condition and hemisphere (error bars indicate the standard deviation). Gray bars, LH; white bars, RH.

DISCUSSION

Mismatch response as an indicator of Voice-specific processing

The present experiment investigated the notion of early specificity in the brain's response to voice information. In a mismatch experiment, subjects were presented with voice signals varying in acoustic similarity and voice prototypicality. Compared to the standard stimulus, each of the deviating voices evoked a mismatch response. In agreement with earlier findings [Giard et al., 1990; Rinne et al., 2000], the SCD modeling revealed a center of the event-related activity in temporal brain regions bilaterally. These results support clinical data reporting a deficit in the discrimination of voices after left or right temporal lobe lesion [Van Lancker and Kreiman, 1987; Van Lancker et al., 1989] as well as fMRI experiments revealing bilateral activation for the perception of human vocalization [Belin et al., 2000]. Moreover, the

temporal lobe regions in non-human primates have also been found to react selectively to species-specific vocalization [Tian et al., 2001].

While prior fMRI experiments on human subjects only investigated the sensitivity for voices compared to nonvocal sounds and frequency-scrambled signals that were not recognizable as voices [Belin et al., 2000], a recent ERP study reported an early characteristic brain response to the auditory presentation of a singing voice compared to instrumental sounds, although the signals were very similar in terms of their acoustic structure [Levy et al., 2001]. In the present experiment, we were able to demonstrate a preattentive sensitivity for prototypicality even within the voice domain. A violation of the listeners' expectations (or perceptual templates) of an average human voice leads to an increase in the mismatch response. This increase in the brain response corresponds inversely to the acceptability rating: the less pleasant a stimulus was rated, the higher the source current density it evoked. The finding that a corresponding outcome was not observed in the "naturalness" rating, is due to the fact that in this rating participants only coarsely distinguished between human speech and non-natural (i.e., manipulated or synthesized) speech whereby any fine-grained differences between the manipulated and the synthetic speech signal were ignored by the raters.

The mismatch responses observed in the present study were remarkably independent of the acoustic signal parameters: Despite its greater acoustic and perceptual similarity to the standard, the F0-manipulated voice (here labeled PITCH) evoked a significantly stronger brain response than the natural female voice. Apparently, the low fundamental frequency in combination with a formant structure typical for female speakers violates the prototypical representation of an average human voice. The mismatch response reflects this violation and is thus interpreted as being dependent on representations of prototypical voices in the auditory long-term store.

A similar dependence on long-term traces has been reported for the language-specific perception of speech sounds. However, when listeners are presented with non-prototypical phoneme deviants, the MMN is smaller than when presented with prototypical phonemes [Näätänen et al., 1997; Winkler et al., 1999]. Since non-prototypical voices contrastingly lead to an increase in the auditory brain response, we argue that the mismatch process to non-prototypical exemplars in the two dimensions (voices, phonemes) are of a different quality. It may be that this difference is due to the different nature of the long-term representations (symbolic for phonemes, nonsymbolic for voices). In addition, it has also been suggested that the two types of information (voice-related , phoneme-related) are segregated into two functionally and neurally different processing streams [Knösche et al., 2002]. Linguistic information is assumed to be processed in a dorsal stream in the posterior part of the auditory cortex and caudal neural pathways. The voice information is assumed to be processed in a ventral stream along the STS [Belin and Zatorre, 2000]. The distinct behavior of the mismatch response to non-prototypical speech sounds vs. non-prototypical voices supports the assumption that the auditory system deals with these types of information in different, specific ways, and shows that this is the case even at preattentive processing stages.

Mismatch response as an indicator of Gestalt-like processing

Most importantly, the increase in the mismatch response observed in the present study has been remarkably independent of acoustic stimulus features. The stronger response to the unusual pitch-shifted stimulus in comparison to the natural female voice is most interesting from the perspective of gestalt-like processing accounts, as it suggests that voice pitch is not analyzed as an independent feature. It rather seems that the pitch of the deviant is not separately compared to the pitch of the standard stimulus independent of the other acoustic parameters. Instead, the mismatch response seems to reflect the gestalt-like processing of voice information, i.e., the mismatch of the pitch and formant configuration. There is an ongoing debate whether the MMN reflects a "gestalt-like" processing of feature conjunctions in complex sounds in general. However, independent of whether the MMN is generally associated with the perception of configurational gestalts [Gomes et al., 1997] or not [Deacon et al., 1998], it appears that the mechanisms involved in voice perception are based on the overall parameter configuration. Similar configurational effects have been reported for face perception [Lewis and Johnston, 1997; Moscovitch et al., 1997; Tanaka and Farah, 1993; Young et al., 1987]. For example, Young et al. [1987] demonstrate, that the discrimination of single features (eyes, chin, etc.) is difficult when they are part of two otherwise different faces [see Hancock et al., 2000]. Furthermore, also in face-recognition, specifically face-sensitive, areas have been localized, namely in the lateral fusiform gyrus,

inferior occipital gyri, and the superior temporal sulcus, and also for the processing of face information, a ventral processing stream has been suggested in temporal lobe areas and the ventral limbic system [Haxby et al., 2000; Kanwisher et al., 1997; Perrett et al., 1984; Sergent et al., 1992]. These observations and the number of further parallels in voice and face processing [Belin et al., 2002; Mann et al., 1979] make it likely that both domains follow common principles.

CONCLUSION

We report neurophysiological evidence for a preattentive sensitivity for human voice information in the brain. Our findings show that ecologically invalid speech signals (computer voices, violating the listener's expectations) lead to a neural response that is stronger than the response to natural speech signals. This finding suggests a mechanism that is voice-specific, i.e., fundamentally different from the processing of non-prototypical speech sounds where a decrease in mismatch-related activity has been reported. The auditory mismatch responses observed in the present study correspond to the overall prototypicality of the presented voice rather than to single acoustic parameters. We, therefore, conclude that voice information is processed not only in a specific, but also in a "gestalt-like" way even at early processing stages.

ACKNOWLEDGMENTS

This work was supported by the Leibniz-Preis dedicated to Angela D. Friederici. We thank E. Schroeger, D. Swinney, and T.C. Gunter for their comments on an earlier version of this paper.

REFERENCES

Belin P, Zatorre R. (2000): 'What', 'where' and 'how' in auditory cortex. Nat Neurosci 3:965–966.

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

Belin P, Zatorre R, Ahad P (2002): Human temporal -lobe response to vocal sounds. Cog Brain Res 13:17–26.

Deacon D, Breton F, Ritter W, Vaughan JH (1998): Automatic change detection: Does the brain use representations of individual stimulus features or gestalts. Psychophysiology 35:413–419.

Fant G (1960): Acoustic theory of speech production. The Hague: Mouton.

Ferguson A, Zhang X, Stroink X (1994): A complete linear discretisation for calculating the magnetic field using the boundary element method. IEEE Biomed Eng 41:459.

Giard M, Perrin F, Pernier J, Boiuchet P (1990): Brain generators implicated in the processing of auditory stimulus deviance; a topographic event-related potential study. Psychophysiology 27: 627–640.

Gomes H, Bernstein R, Ritter W Jr, Vaughan H, Miller J (1997): Storage of feature conjunctions in transient auditory memory. Psychophysiology 34:712–716.

Hämälainen M, Ilmoniemi R (1994): Interpreting magnetic field of the brain: minimum norm estimates. Med Biol Comp 32:183.

Hancock P, Bruce V, Burton A (2000): Recognition of unfamiliar faces. Trends Cogn Sci 4:330–337.

Hari R, Aittoniemi K, Jarvinen M, Katila T, Varpula T (1980): Auditory evoked transient and sustained magnetic fields of the human brain: Localization of neural generators. Exp Brain Res 40:237–240.

Haxby JV, Hoffman E, Gobbini M (2000): The distributed human neural system for face perception. Trends Cogn Sci 4:223–233.

Kanwisher N, McDermott J, Chun M (1997): The fusiform face area: a module in human extrastriate cortex specialized for face perception. J Neurosci 17:4302–4311.

Knösche T., Lattner S., Maess B., Schauer M, Friederici A (2002): Early parallel processing of auditory word and voice information. NeuroImage 17:493–1503.

Levy DD, Granot R, Bentin S (2001): Processing specificity for human voice stimuli: electrophysiological evidence. NeuroReport 12:2653–2657.

Lewis M, Johnston R (1997): The Thatcher-Illusion as a test of configurational disruption. Perception 26:225–227.

Mann V, Diamond R, Carey S (1979): Development of voice recognition: parallels with face recognition. J Exp Child Psychol 27: 153–165.

Moscovitch M, Wincour G, Behrmann M (1997): What is special about face recognition? Nineteen experiments on a person with visual object agnosia and dyslexia but normal face recognition. J Cogn Neurosci 9:555–604.

Näätänen R (1992): Attention and brain function. Hillsdale: Lawrence Erlbaum Associates.

Näätänen R, Gaillard A, Mäntysalo S (1978): Early selective-attention effect on evoked potential reinterpreted. Acta Psychol 42: 313–329.

Näätänen R, Schröger E, Karkas S, Tervaniemi M, Paaviainen P (1993): Development of a memory trace for a complex sound in the human brain. NeuroReport 4:503–506.

Näätänen R, Lehtokoski A, Lennes M, Cheour M, Houtilainen M, Ilvonen A, et al. (1997): Language-specific phoneme representations revealed by electric and magnetic brain responses. Nature 385:432–434.

Näätänen R, Tervaniemi M, Sussman E, Paavilainen P, Winkler I (2001): Primitive intelligence in the auditory cortex. Trends Neurosci 24:283–287.

Nusbaum H, Francis A, Henly A (1995): Measuring the naturalness of synthetic speech. Int J Speech Technol 1:7–19.

Perrett D, Smith P, Potter D, Mistlin A, Head A, Milner A, et al. (1984): Neurones responsive to faces in the temporal cortex: studies of functional organization, sensitivity to identity and relation to perception. Hum Neurobiol 3:197–208.

Rinne T, Alho K, Ilmoniemi R, Virtanen J, Näätänen R (2000): Separate time behaviors of the temporal and frontal mismatch negativity sources. NeuroReport 10:1113–1117.

Sams M, Paavilainen P, Alho K, Näätänen R (1985): Auditory Frequency discrimination and event-related potentials. Electroenceph Clin Neurophysiol 62:437–448.

Sergent J, Ohrta S, MacDonald B (1992): Functional neuroanatomy of face and object processing. A positron emission tomography study. Brain 115:15–36.

Sinkkonen J, Tervaniemi M (2000): Towards optimal recording and analysis of the mismatch negativity. Audiol Neurootol 5:235–246.

Tanaka J, Farah M (1993): Parts and wholes in face recognition. Q J Exp Psychol 46A:225–245.

Tian B, Teser D, Durham A, Kustov A, Rauschecker J (2001): Functional specialization in rhesus monkey auditory cortex. Science 313:290–293.

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

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

Van Lancker D, Kreiman J, Cummings J (1989): Voice perception deficits: neuroanatomical correlates of phonagnosia. J Clin Exp Neuropsych 11:665–674.

Vihla M, Lousnasmaa O, Salmelin R. (2000): Cortical processing of change detection: Dissociation between natural vowels and two-frequency complex tones. Proc Natl Acad Sci 97:10590–10594.

Wang J, Williamson S, Kaufman L. (1992): Magnetic source images determined by a lead-field analysis: the unique minimum-norm least-squares estimation. IEEE Biomed Eng 39:665–675.

Winkler I, Kujala T, Tiitinen H, Sivonen P, Alku P, et al. (1999): Brain responses reveal the learning of foreign language phonemes. Psychophysiology 36:638–642.

Young A, Hellawell D, Hay DC. (1987): Configurational information in face perception. Perception 16:747–759.