Now You Hear It, Now You Don't: Transient Traces of Consonants and their Nonspeech Analogues in the Human Brain

Jonas Obleser[1,2], Sophie K. Scott[1] and Carsten Eulitz[2]

[1]Institute of Cognitive Neuroscience, University College London, 17 Queen Square, London, WC1N 3AR, UK and [2]University of Konstanz, PO Box D25, 78457 Konstanz, Germany

The apparently effortless identification of speech is one of the human auditory cortex' finest and least understood functions. This is partly due to difficulties to tease apart effects of acoustic and phonetic attributes of speech sounds. Here we present evidence from magnetic source imaging that the auditory cortex represents speech sounds (such as [g] and [t]) in a topographically orderly fashion that is based on phonetic features. Moreover, this mapping is dependent on intelligibility. Only when consonants are identifiable as members of a native speech sound category is topographical spreading out in the auditory cortex observed. Feature separation in the cortex also varies with a listener's ability to tell these easy-to-confuse consonants from one another. This is the first demonstration that speech-specific maps of features can be identified in human auditory cortex, and it will further help us to delineate speech processing pathways based on models from functional neuroimaging and non-human primates.

Keywords: auditory cortex, consonants, intelligibility, magnetic source imaging, magnetoencephalography, MEG, N100, N100m speech

Introduction

Whenever our ear is hit by speech, a cascade of automatic processing steps takes place, leading to a surprisingly robust mapping of the heard sound stream onto meaning. This is possible only because of efficient, yet largely unrevealed decoding of the speech signal throughout structures of the auditory pathway. Within recent years and mainly due to the emergence of powerful neuroimaging techniques, the neuroanatomical structures subserving speech perception have been unravelled. It has been shown repeatedly and is widely accepted that structures surrounding the primary auditory areas (located on medial parts of Heschl's gyrus in the supratemporal plane) are crucially involved in speech processing. Typically, the anterior and lateral parts of the superior temporal gyrus and sulcus and, less consistently, the inferior frontal gyrus are activated more vigorously by speech sounds than by nonspeech noise or pseudo speech matched in acoustic complexity (such as spectrally inverted speech; Binder et al., 2000; Zielinski and Rauschecker, 2000; Scott et al., 2000; Zatorre et al., 2002; Narain et al., 2003; Callan et al., 2004; Obleser et al., 2005; Liebenthal et al., 2005). Although the results obtained with functional magnetic resonance imaging (fMRI) and positron emission tomography (PET) provide structural details of functional neuroanatomy, the severe temporal insensitivity of the method may pose a problem when the processing of short, transient speech signals is studied. Magnetoencephalography (MEG), with its superior temporal resolution, allows the precise delineation of the temporal processing of such transient stimuli. MEG also has an acceptable spatial resolution, especially for

tangential generators (such as cortical tissue in the supratemporal plane; Pantev et al., 1995; Fujioka et al., 2002; Lütkenhöner et al., 2003), and it also allows powerful relative comparisons of conditions within subjects (Hämäläinen et al., 1993; Lounasmaa et al., 1996). A major advantage for speech processing is the potential to identify the signature of an initial processing step in the speech sound decoding cascade, e.g. as reflected in the N100/N100m brain wave deflection elicited ~100 ms after stimulus onset by a vast array of auditory events in virtually every healthy subject (Näätänen and Picton, 1987; Näätänen and Winkler, 1999).

Studies recording the N100m response to natural vowels (Mäkelä et al., 2003; Obleser et al., 2003a, 2004a,b; Shestakova et al., 2004) and syllables (Obleser et al., 2003b) and its approximate source in auditory cortex have indicated that a topographical mapping of feature dimensions in auditory cortex might play a supporting role in speech sound perception. Thus neural networks along the auditory central pathway may exploit the distributional properties inherent to the acoustics of incoming speech sounds (Jusczyk, 2002; Maye et al., 2002; Pena et al., 2002), leading to a partial topographical separation for consistently uncorrelated input and to an overlapping representation for common features in the input (Buonomano and Merzenich, 1998; Kohonen and Hari, 1999). The commonalities among speech sounds can be quantified, since vowels and consonants, although vastly different in spectro-temporal shape, share certain features in articulation (Chomsky and Halle, 1968). It is these features, such as place of articulation (back of the tongue, termed back, velar, dorsal or tip/body of the tongue, referred to as front, alveolar/dental, coronal) that crucially influence the acoustic information available to be processed and utilized in human speech perception. However, whereas vowels exhibit a steady-state pattern of characteristic spectral peaks (formants) that are perceptually relevant to vowel identification (Peterson and Barney, 1952; Hose et al., 1983), stop consonants are temporally much more transient and acoustically much more variable. Nevertheless, stop identification works surprisingly well in running speech.

Here, we want to test whether the topographical mapping of the acoustic consequences of different place features repeatedly seen in vowels also holds for isolated, highly transient stop consonantal bursts. A 50 ms stop burst edited from natural utterances of various speakers and contexts is intelligible, i.e. it can be accordingly categorized by the listener. Nevertheless, due to its acoustics, it delineates the border between speech perception and mere perception of a complex acoustic non-speech sound. This study is set to test the speech specificity of the transient processing step reflected in the N100m by comparing responses to natural consonants with responses to

their acoustically as complex, yet unintelligible (spectrally inverted; Blesser, 1972) analogues. 'Intelligibility' here refers to the comprehensibility of a consonant, i.e. a fully intelligible consonant could be understood and repeated by a skilled speaker of the relevant language. If the N100m topographical mapping indicates such speech-specific perception to a certain degree, spatial arrangement of N100m generators in response to different consonants should be affected by intelligibility. Additionally, we will be able to exploit the temporal sensitivity of MEG to compare different stages of processing in the aftermath of a short consonantal burst, an analysis currently not possible with functional MRI.

Materials and Methods

Subjects
Nineteen volunteers (24.3 ± 2.4 years, mean ± SD) took part in this experiment. All subjects were monolingual native Germans, right-handed (score > 90) according to the Edinburgh handedness test (Oldfield, 1971), and reported no history of neurological or otological disorders. They signed an informed consent form and were reimbursed with €15.

Stimuli
The stimulus set consisted of 128 different speech and non-speech sounds. Four different stops (front-voiced [d], front-unvoiced [t], back-voiced [g], back-unvoiced [k]; Table 1), with four different ensuing vowel contexts in the words from which they were edited ([i:], [e:], [u:], [o:]), and spoken by four different speakers (two males, two females). Each of these 64 items was used twice, in an intelligible format as well as in an unintelligible, spectrally inverted format (see below).

Original recordings were accomplished using a DAT-recorder (sampling rate 48 kHz) and a high-quality microphone in a soundproof chamber. The stimulus material was redigitized with a 20 kHz sampling rate (mono, 16 bit resolution) and edited off-line using SoundForge™ (Sonic Foundry) and Matlab™ (Mathworks). Resulting files were 50 ms long, starting at the last zero crossing before onset of the stop burst (pre-voicing was left unaltered if present), and faded out with a 10 ms Gaussian ramp.

From each consonant audio file, two exemplars to be used for stimulation were created. One unintelligible exemplar was derived by applying the spectral inversion procedure by Blesser (1972), as used in brain imaging studies previously (Scott et al., 2000; Narain et al., 2003). Another intelligible exemplar was maintained by simply applying a zero-phase Butterworth lowpass filter (cut-off 4 kHz). Most relevant to an auditory evoked-field study, the envelopes of both sounds do not differ, and the spectro-temporal complexity of the consonant is preserved (Fig. 1). However, the typical spectral distribution is changed, rendering the speech sound unintelligible. The frequency spectrum is flipped at ~2 kHz, e.g. turning a typical spectral peak in a [d] (>2.5 kHz) into a peak of <1.5 kHz. We attempted to match the long-term power spectra of original and inverted signals by applying a pre-emphasis high-pass filter prior to spectral inversion (Scott et al., 2000). All final

stimulus files were matched to equal loudness (i.e. root mean squared amplitude was adjusted).

Stimulation and Behavioural Testing
First, hearing threshold of the right ear was determined individually using a randomly selected exemplar of [g] (in pilot testing in several subjects, [g] had consistently elicited most insensitive thresholds, and no further consistent threshold differences between stimuli were observed). Stimulation loudness was then adjusted to 55 dB above the respective threshold.

Stimuli were presented monaurally to the right ear using Presentation™ (Neurobehavioral Systems) and a customized sound delivery system with 6-m long air-conduction tubes and plastic in-ear pieces as headphone substitutes (approximately linear frequency transmission between 0.2 and 4 kHz).

In a behavioural pre-test, subjects were presented a pseudo-randomized sequence of 32 intelligible and 32 unintelligible stimuli with a fixed onset asynchrony of 3 s (total duration ~3 min), and they had to categorize the heard sounds on a multiple-choice five-category list ([d], [t], [g], [k] and none; the first two items were excluded from score calculation). For prior successful application of speech sound fragment categorization tasks, see Obleser et al. (2003b).

In the actual MEG measurements, subjects listened attentively without any further task or distraction to a randomized sequence of 1152 stimuli, presented with a randomized onset asynchrony of 1.8–2.2 s between stimuli. Recording was interrupted twice by short subject-paced breaks.

MEG Recording and Data Analysis
Auditory magnetic fields were recorded using a 148-channel whole head neuromagnetometer (Magnes 2500, 4D Neuroimaging) in a magnetically shielded room (Vaccumschmelze). Epochs of 500 ms duration (including a 100-ms pre-trigger baseline) were recorded with a bandwidth from 0.1 to 200 Hz and a 687.17 Hz sampling rate. If the peak-to-peak amplitude exceeded 3.5 pT in one of the channels or the co-registered electrooculogram (EOG) signal was larger than 100 μV, epochs were rejected.

We analysed up to 144 artefact-free consonant responses that remained for each consonant category after offline noise correction, and averaged them separately for consonant category ([d],[t],[g],[k]) and intelligibility but across speaker voice and vowel context. The resulting averages thus contained brain responses to eight acoustically variant exemplars of a consonant (or an unintelligible version thereof), which makes results more comparable to our previous studies. A 20 Hz low-pass filter (Butterworth 12 dB/oct, zero phase shift) was subsequently applied to the averages.

The N100m component was evident in all subjects and all conditions and was defined as the prominent waveform deflection in the time range between 90 and 160 ms. Isofield contour plots of the magnetic field distribution were visually inspected to ensure that N100m and not P50m or P200m was analysed. N100m peak latency was defined as the sampling point in this latency range by which the first derivative of the root mean square amplitude reached its minimum and second derivative was smaller than zero. Root mean square amplitude was calculated across 34 magnetometer channels selected to include the field extrema over the left hemisphere.

Prior to statistical analyses, all brain response latencies were corrected for a constant sound conductance delay of 19 ms in the delivery system. Using the same set of channels, an equivalent current dipole (ECD) in a spherical volume conductor (fitted to the shape of the regional head surface) was modelled at every sampling point (Sarvas, 1987). The resulting ECD solution represents the centre of gravity for the massed and synchronized neuronal activity. An ECD solution was considered anatomically plausible if its location was >2 cm in a medial-lateral direction from the centre of the brain and ~3–8 cm in superior direction, measured from the connecting line of the pre-auricular points. As source location displacements do not appear exactly and exclusively along the Cartesian axes of the source space (Braun et al., 2002), we additionally measured N100m ECD location in polar angle Φ and azimuth angle θ, which quantify angular displacements in the sagittal and the axial plane, respectively.

Table 1
Overview of the acoustic stimuli used, their features and their acoustic characteristics (mean and SD of centre frequency of the consonantal burst as determined with linear predictive coding, LPC)

	Consonant	Voicing	Place	No. of items	Mean frequency of burst (Hz)	(SD)
Intelligible	[d]	Voiced	Front	16	2600	(353)
	[g]	Voiced	Back	16	2448	(344)
	[t]	Unvoiced	Front	16	2824	(190)
	[k]	Unvoiced	Back	16	2475	(465)
Unintelligible (spectrally inverted)	[d]	Voiced	Front	16	2196	(275)
	[g]	Voiced	Back	16	2170	(269)
	[t]	Unvoiced	Front	16	1860	(157)
	[k]	Unvoiced	Back	16	2144	(420)
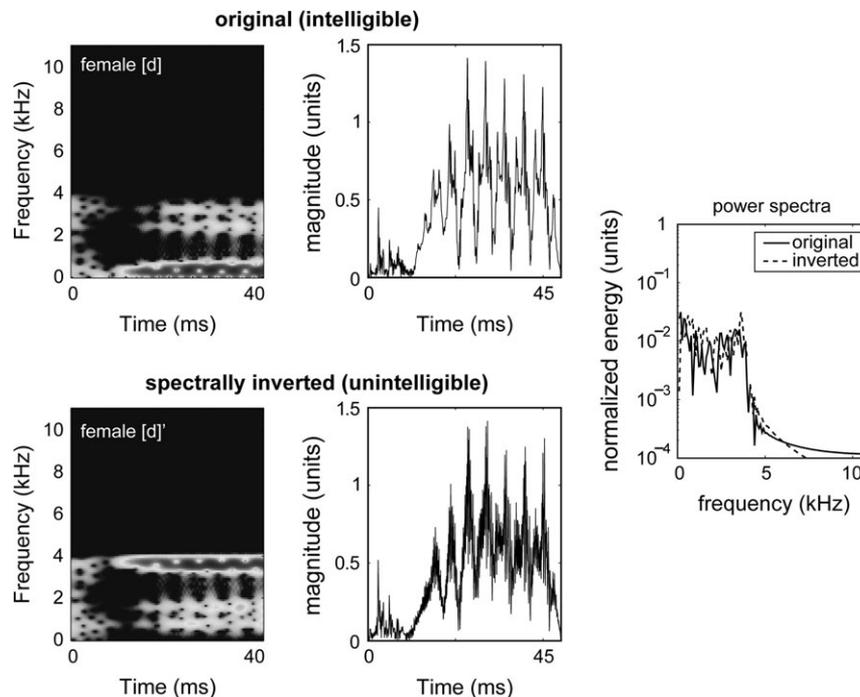
Figure 1. Illustration of the effect of spectral inversion. It renders a stop consonant [d] (upper panels) into an unintelligible version [d]′ (lower panels), while preserving the spectro-temporal complexity (cf. spectrograms, left) and the temporal envelope structure (cf. magnitude of the analytic signal, middle panels). Differences in power spectrum (rightmost panel) cannot be eliminated entirely and are inherent to the method of spectral inversion.

Statistical Analysis

Influences of the consonants on N100m latency and amplitude as well as on three-dimensional location of the N100m dipole solution were tested with a repeated measures analysis of variance (Mixed Model, SAS™, *SAS Institutes*) with fixed factors Place of articulation (front, back), Voicing (voiced, unvoiced) and Intelligibility (intelligible, unintelligible) and the random factor Subject. [Most relevant to repeated measures designs, the outstanding advantage of using general linear mixed models with a restricted maximum likelihood estimate instead of a least squares estimate is that single observations from subjects can be discarded (e.g. due to invalid ECD solutions) without losing the subject's data entirely: other valid observations from the subject will be used in the model nevertheless and will contribute to the enhanced power of a larger subject sample (Wolfinger, 1997; Carrier *et al.*, 2001; Gandour *et al.*, 2004; Frost *et al.*, 2004).] Vowel context and speaker were balanced but not studied as factors here, and the analysis focused on the left hemisphere (contralateral to the stimulated ear). We also tested possible influences on the ensuing auditory evoked field by calculating and analysing the median ECD solution over 250–300 ms (late field) and over 300–350 ms (very late field), respectively. The time slices were selected by inspecting the sustained field of the grand mean root mean square waveforms, which indicated two additional peaks approximately in the 250–300 ms and the 300–350 ms ranges across conditions. For a previous application of median ECD solutions representing sustained-field activity, see Eulitz *et al.* (1995).

Results

Behavioural data

Mean percentage correct in a forced-choice categorization task of the stop consonants [d], [t], [g] and [k] (plus a category for non-speech) amounted to 57.2 ± 18.3% (mean ± SD). With chance level at 20% due to five response categories, all subjects performed above chance. Errors were evenly distributed across actual speech exemplars (67.4 ± 18.8% correct speech item categorization) and unintelligible exemplars (67.4 ± 27.3% correct non-speech item categorization). Most relevant, error rates specific to certain phonological features were low; voicing categorization was correct in 87.6 ± 13% and place categorization in 88.5 ± 11% of all viable items, respectively.

Brain Responses

A very prominent N100m response was detectable over the left hemisphere in all subjects and conditions (Figs 2 and 3). The mean signal-to-noise ratio was 9.1 (± 4) to 1 (mean ± SD; N100m peak root mean square amplitude divided by mean baseline root mean square amplitude; intelligible 9.7:1, unintelligible 8.5:1). Mean goodness of fit between the single ECD model and the measured magnetic field distribution amounted to 0.96 ± 0.04 (intelligible 0.95, unintelligible 0.96), and mean confidence volume indicating 95% certainty of source location was 522 ± 1466 mm³ (intelligible 621, unintelligible 423). N100m morphology as expressed by peak latency and amplitude revealed the following: Peak latency was affected independently by the acoustic consequences of features Place [$F(1,102) = 5.83$, $P < 0.02$] and Voicing [$F(1,102) = 35.82$, $P < 0.0001$]. Responses to fronted consonants [d] and [t] peaked ~7 ms earlier than responses to back consonants [g] and [k], and voiced consonants [d] and [g] elicited peaks ~13 ms delayed compared to voiceless consonants [t] and [k]. Both effects were independent of intelligibility. Peak amplitude was affected by interacting factors Voicing and Intelligibility [$F(1,102) = 5.73$, $P < 0.02$]; intelligible voiced consonants yielded by 39 fT stronger N100m amplitudes than their unvoiced counterparts, an effect that diminished to ~9 fT in unintelligible items.

Highly congruent to previous studies, a topographical mapping of consonantal feature Place could be observed on the N100m source. However, this mapping clearly depended on Intelligibility [$F(1,102) = 7.02$, $P < 0.01$]. Among intelligible
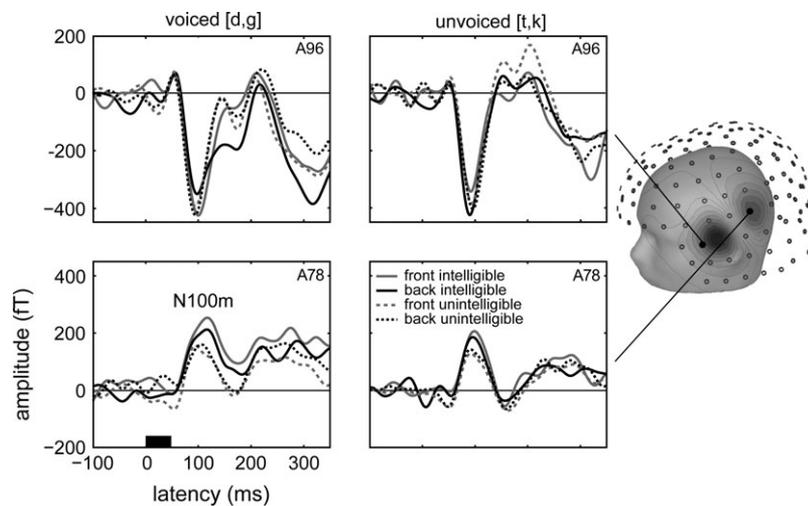
Figure 2. A single subject's auditory evoked fields in response to intelligible (solid) and unintelligible (dashed) consonants. Left panels show responses to voiced items [d] (grey) and [g] (black), unvoiced [t] (grey) and [k] (black) are shown on the right. Upper panels show data from the maximally responsive left-anterior channel, lower panels show data from maximally responsive left-posterior channel (see rightmost reference display). Black bar in the lower left panel indicates stimulus duration.

speech sounds, the N100m in response to the consonants [d] and [t] (front) originated almost 7 mm more anterior in left supratemporal gyrus than the response to [g] and [k] (back, $P < 0.03$). This effect vanished entirely for the unintelligible counterparts (Fig. 3). The spreading of responses was blurred, the variability of the N100m source location across subjects to unintelligible front and back consonants was considerably large, and the spatial difference did not attain significance.

This interaction of consonantal feature Place and Intelligibility was not restricted to the posterior–anterior dimension and can be described more accurately as an angular shift of N100m sources in the axial [Place × Intelligibility interaction, $F(1,102) = 5.78$, $P < 0.02$] and sagittal planes [$F(1,102) = 6.21$, $P < 0.02$], with the responses to intelligible front consonants [d] and [t] originating from most anterior, most medial and most superior locations (Fig. 4).

To test for any functional significance of this spatial spreading of consonant features in the N100m brain mapping, the correlation with subject's ability to discriminate stops of different place of articulation correctly (see behavioural results above) was calculated. Most intriguingly, the spreading in the posterior–anterior dimension in response to intelligible speech sounds was a linear function of subject's correct place identification in the behavioural task ($r = 0.46$, $P < 0.02$, explaining 21.1% of variance) (Fig. 5a). As expected, this did not hold for brain responses to unintelligible sounds ($r = 0.14$, $P < 0.45$). In contrast, there was a trend for responses to non-speech sounds to be more spread out in those subjects committing more speech/non-speech classification errors in general ($r = -0.24$, $P < 0.20$, explaining 6% of variance).

Transformation of grand mean head coordinates to Talairach space revealed that N100m activity in response to these consonantal bursts most likely emerged from the more lateral section of Heschl's gyrus (BA 42, Talairach coordinates x, y, $z = [-59, -27, 9]$; Fig. 4).

Interestingly, the effect of topographical mapping was only evident during the N100m time window. Activation during late and very late time windows did not show such Place × Intelligibility interaction whatsoever (both $F < 1$). Notwithstanding, a difference in amplitude (dipole moment $|Q|$,
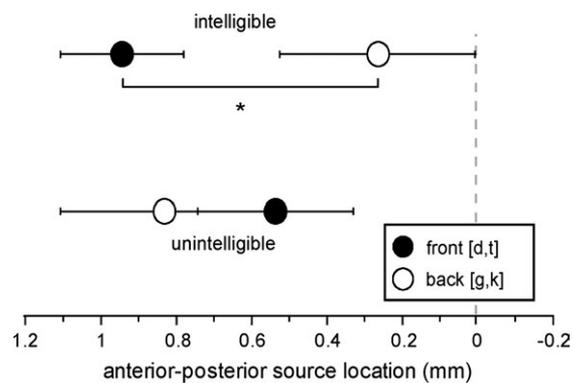

Figure 3. The mapping of consonants differing in place of articulation (front consonants [d, t], black dot, versus back consonants [g, k], white dot) along the posterior–anterior dimension in auditory cortex is affected by intelligibility: whereas the mapping is spread out in intelligible consonants (upper dots, *$P < 0.03$), it is blurred when unintelligible consonants are presented (lower dots, error bars indicating SD overlap extensively).

approximating the number of neurons synchronously activated) not evident on the N100m appeared; in the 250–300 ms time window, voiced consonants irrespective of intelligibility led to 6 nAm stronger brain responses [Voicing, $F(1,65) = 5.43$, $P < 0.03$]. This developed into a Voicing × Intelligibility interaction in the 300–250 ms time window [$F(1,61) = 7.93$, $P < 0.01$]. [d] and [g] (29 nAm) elicited a more vigorous response than [t] and [k] (17 nAm) among intelligible items ($P < 0.0001$), which virtually regressed to the mean among spectrally inverted, unintelligible items (22 nAm in both conditions, $P > 0.50$).

To ensure that the topographical mapping of intelligible front and back consonants was reliable across the subject sample and not due to a sampling artefact, a jackknifing procedure was applied. The Place × Intelligibility interaction test on posterior–anterior N100m location was applied iteratively 19 times, with one of the 19 subjects omitted at a time. If the effect were driven by a single outstanding subject, the F-value should drop dramatically when this subject is omitted. Figure 5b shows that this is not the case; thus, the interaction appears to reflect
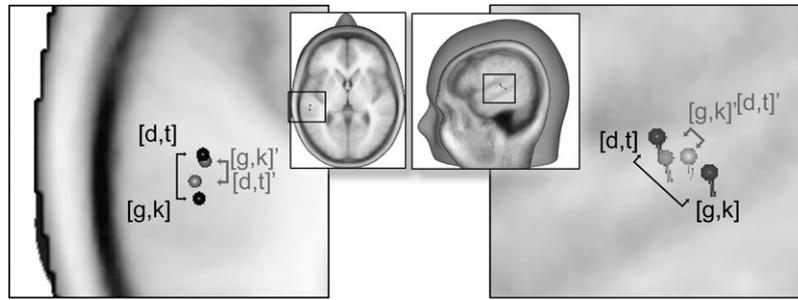
Figure 4. Overlay of grand mean N100m source locations for front and back consonants in intelligible versions (shown in black, with phonetic symbols [d, t] and [g k]) and for the unintelligible versions (shown in grey, denoted with [d, t]′ and [g, k]′ accordingly) onto a standard MR brain template. Left panel depicts an axial view, right panel a sagittal view (zoomed areas are indicated on overview slices). Note in both views the difference in distance between centres of gravity for the intelligible versus unintelligible consonant responses (black symbols versus grey symbols).
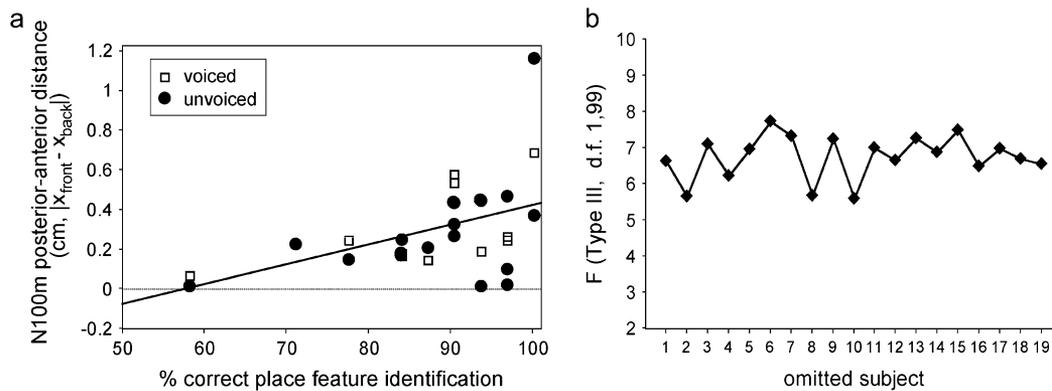
Figure 5. (*a*) Correlation of the reported N100m feature spreading along the posterior–anterior dimension for intelligible speech sounds (ordinate, each subject contributing a within-voiced and a within-unvoiced measure) with subject's performance of place feature discrimination (errors on [d, t] versus [g, k] classification divided by all intelligible items in the task, absicca). Open squares indicate voiced [d–g] N100m distance, black dots the unvoiced [t–k] distance. (*b*) Demonstration of internal consistency for the Place of articulation × Intelligibility interaction. *F*-values (black diamonds) with one subject omitted at a time (abcissa) are displayed. Although some changes in *F*-value are observed, the overall significance level is $P < 0.02$ in all instances.

a general processing property of the auditory system manifest in the N100m time window.

Discussion

This study set out to scrutinize the temporal and topographical mapping of consonantal speech sounds in human auditory cortex, and to determine the influence of intelligibility on this mapping, as opposed to purely acoustic variation. Would acoustic manipulations that preserve the spectro-temporal complexity while rendering the consonants unintelligible evoke the same N100m mapping? If so, the N100m could hardly be interpreted as reflecting a cognitive processing step, such as feature integration and categorical perception.

Previous studies indicate that there are top-down influences onto the N100/N100m process (Näätänen and Picton, 1987; Näätänen and Winkler, 1999; Sanders *et al.*, 2002; Obleser *et al.*, 2004a). The current data add to this evidence, as the N100m response to isolated stop consonants and its origin in auditory cortex were clearly affected by the intelligibility of these speech sounds. For regular intelligible speech sounds, we identified a spreading of the N100m locations within the sagittal plane, as we (Obleser *et al.*, 2003b, 2004a,b) and others (Mäkelä *et al.*, 2003) reported previously for vowels and consonant–vowel syllables. Also highly consistent with our previous results, the

mapping was driven by the acoustic consequences of the place of articulation feature, i.e. [d] and [t] sources were located more anterior than [g] and [k] sources (Fig. 4), irrespective of the voicing of these stimuli. Most relevant to the functional significance of this mapping, the feature-driven difference was not evident among unintelligible consonant analogues (Figs 3 and 4). Whereas isolated intelligible consonants elicited centres of gravity almost 7 mm apart, unintelligible analogues activated clusters separated only by 3 mm, with highly overlapping variance.

This is the first demonstration of an intelligibility or distinctiveness effect of a speech sound set onto the extent of an N100m mapping. In the unaltered intelligible stimulus set, single exemplars invariably tapped onto learned categorical percepts on the one hand and the brain response was spread out topographically depending on salient feature place of articulation on the other hand. In the altered stimulus set (matched for spectro-temporal complexity and diversity), single exemplars obviously did not tap onto acquired phoneme categories and concomitantly no topographical spreading out of N100m responses was observed. Even more so, the actual extent of the N100m feature map was correlated with subjects' ability to correctly categorize the intelligible stop consonants' place of articulation. Subjects making more [d]/[g] or [t]/[k] errors in this comparably difficult task tended to exhibit a less

spread-out feature map in the subsequent brain recordings. Such an interdependence of performance and spatial extent of N100m mapping has been implied previously in congenitally blind subjects, depending exceptionally on their auditory perceptual system (Elbert *et al.*, 2002). One might therefore expect to find a mild positive correlation between perceptual performance on the behavioural level and extent of a cortical map of the relevant features, here reflected in the N100m source locations.

In a previous study using combinations of voiced stop consonants and vowels, the N100m mapping was found to be driven almost exclusively by features of the vowel (Obleser *et al.*, 2003b). This result was interpreted as a dominance of vowel perception processes being reflected in the N100m time range, the vowel being also the articulatory target state with much lower acoustic variability than the preceding stop consonant (Fitch *et al.*, 1997). However, another possible source for this vowel preponderance is a potential backward masking of more consonant-specific processes through processing of the ensuing steady-state vowel portion (Koyama *et al.*, 2003). This was one of the main reasons why we chose the present isolated-stop design. The result, namely a consonant feature map evident in the N100m source locations that highly resembles the previously identified vowel feature maps, justifies this effort and adds evidence to a ubiquitous feature mapping that is abstract of simple acoustic determinants.

Although source activation differences between intelligible and unintelligible speech sounds was evident in the aftermath (most likely reflecting different processing depth of intelligible, hence classifiable items and nonsense items), the interesting interaction of intelligibility and phonological features vanished in later stages. Since processing of meaningful (i.e. intelligible) sounds is usually stronger in lateral and anterior sections of the superior temporal cortex (Scott *et al.*, 2000; Narain *et al.*, 2003; Liebenthal *et al.*, 2005), the feature mapping seen comparably early in time (~100 ms post-stimulus onset) and hierarchically rather low in functional neuroanatomy (lateral bank of Heschl's gyrus instead of planum polare or superior temporal sulcus) could reflect a necessary precursory step in speech processing. The highly pre-processed acoustic information has to be reintegrated and abstracted to an invariant pattern to form a unitary percept of speech, and N100m responses might partly reflect this mechanism (Näätänen and Winkler, 1999; Krumbholz *et al.*, 2003). The output of such percept-forming processing stages may then allow for the higher-level areas in anterior temporal cortex and what has been termed the auditory 'what' system (Rauschecker and Tian, 2000; Scott and Johnsrude, 2003) to process meaning and content of speech, as shown in neuroimaging studies (Scott *et al.*, 2000; Narain *et al.*, 2003; Liebenthal *et al.*, 2005). However, the precursory transient processing stages of percept formation and evaluation might be missed out when using time-insensitive measures such as functional MRI and PET. Additionally, the 50 ms single consonant bursts were very suitable stimuli to elicit a vigorous auditory evoked response in all our subjects, whereas BOLD signal change in response to such highly transient stimuli can be expected to be fairly low (Robson *et al.*, 1998; Tanaka *et al.*, 2000).

Magnetic source imaging is known to be more reliable for the relative comparisons between sources (which constitute the main result we reported above) than for its absolute spatial certainty (Lounasmaa *et al.*, 1996). Surprisingly, however, the mean centre of activation we report above for the N100m source location deviates from the mean site of disrupted stop consonant perception during electrocortical stimulation mapping (Boatman and Miglioretti, 2005) by only 1.1 mm in the sagittal plane (5.6 mm in the axial plane, with the N100m site being located slightly more medial than the centre of disruption during electrocortical stimulation). This implies that N100m source imaging and invasive electrocortical stimulation at least partly tap onto the same perceptual processes.

No conclusion upon hemispheric differences can be drawn since only the contralateral left-hemispheric response to monaurally presented stimuli was studied. It should be noted, however, that evidence for strongly left-lateralized speech processing of phonemes and syllables in the N100m is sparse (Obleser *et al.*, 2003a,b, 2004b; but see Mäkelä *et al.*, 2003).

Interestingly, the N100m also exhibited peak latency differences due to mainly spectral (Place, spectral peak of [d,t] versus [g,k]) and temporal (voicing, voice onset time, [d,g] versus [t,k]) features of the stimuli. This corroborates previous findings of auditory evoked potential/field peak latency and its role in speech sound processing (Gage *et al.*, 2002; Roberts and Gage, 2002; Eulitz and Lahiri, 2004; Obleser *et al.*, 2004b). However, these effects were not modulated by intelligibility. With respect to the mainly temporal feature voicing and looking at the effect of spectral inversion on the stimulus time course (Fig. 1), this is not much of a surprise. However, latency effects of the feature having mainly spectral consequences, place of articulation, may have been expected to be reduced or blurred when unintelligible analogues are presented — which they were not. This is a strong hint that topographical mapping (affected by intelligibility) and temporal mapping (unaffected) signify distinct contributions to the perception of speech sounds in the N100m latency range, as proposed previously (Obleser *et al.*, 2004b).

Using natural stop consonants and spectrally inverted analogues, we have reported an influence of speech intelligibility on the topographical mapping of a ubiquitous speech sound feature, place of articulation, and its acoustic consequences in human auditory cortex. We demonstrated that this mapping is independent of exact acoustic realization and also of other features such as voicing. However, intelligibility in the sense of one's ability to classify an incoming signal and map it onto a learned speech sound category is necessary to activate the feature mapping reported previously in vowels and consonant–vowel syllables. Moreover, we showed that the feature mapping appears to bear a direct connection to subject's ability to correctly classify the place feature when listening to isolated stop consonants. The results (i) show that highly time-sensitive measures with a reasonable spatial resolution such as auditory evoked fields have a role to play when studying the elusive speech signal; and (ii) form another building block in our understanding of the human speech faculty and its neurobiological underpinnings.

Notes

Address correspondence to Dr Jonas Obleser, Institute of Cognitive Neuroscience, 17 Queen Square, London WC1N 3AR, UK Email: jonas@obleser.de.

References

Binder JR, Frost JA, Hammeke TA, Bellgowan PS, Springer JA, Kaufman JN, Possing ET (2000) Human temporal lobe activation by speech and nonspeech sounds. Cereb Cortex 10:512-528.

Blesser B (1972) Speech perception under conditions of spectral transformation. I. Phonetic characteristics. J Speech Hear Res 15:5-41.

Boatman DF, Miglioretti DL (2005) Cortical sites critical for speech discrimination in normal and impaired listeners. J Neurosci 25:5475-5480.

Braun C, Haug M, Wiech K, Birbaumer N, Elbert T, Roberts LE (2002) Functional organization of primary somatosensory cortex depends on the focus of attention. Neuroimage 17:1451-1458.

Buonomano DV, Merzenich MM (1998) Cortical plasticity: from synapses to maps. Annu Rev Neurosci 21:149-186.

Callan DE, Jones JA, Callan AM, Akahane-Yamada R (2004) Phonetic perceptual identification by native- and second-language speakers differentially activates brain regions involved with acoustic phonetic processing and those involved with articulatory-auditory/orosensory internal models. Neuroimage 22:1182-1194.

Carrier J, Land S, Buysse DJ, Kupfer DJ, Monk TH (2001) The effects of age and gender on sleep EEG power spectral density in the middle years of life (ages 20-60 years old). Psychophysiology 38:232-242.

Chomsky N, Halle M (1968) The sound pattern of English. New York: Harper & Row.

Elbert T, Sterr A, Rockstroh B, Pantev C, Muller MM, Taub E (2002) Expansion of the tonotopic area in the auditory cortex of the blind. J Neurosci 22:9941-9944.

Eulitz C, Lahiri A (2004) Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition. J Cogn Neurosci 16:577-583.

Eulitz C, Diesch E, Pantev C, Hampson S, Elbert T (1995) Magnetic and electric brain activity evoked by the processing of tone and vowel stimuli. J Neurosci 15:2748-2755.

Fitch RH, Miller S, Tallal P (1997) Neurobiology of speech perception. Annu Rev Neurosci 20:331-353.

Frost C, Kenward MG, Fox NC (2004) The analysis of repeated 'direct' measures of change illustrated with an application in longitudinal imaging. Stat Med 23:3275-3286.

Fujioka T, Kakigi R, Gunji A, Takeshima Y (2002) The auditory evoked magnetic fields to very high frequency tones. Neuroscience 112:367-381

Gage NM, Roberts TP, Hickok G (2002) Hemispheric asymmetries in auditory evoked neuromagnetic fields in response to place of articulation contrasts. Brain Res Cogn Brain Res 14:303-306.

Gandour J, Tong Y, Wong D, Talavage T, Dzemidzic M, Xu Y, Li X, Lowe M (2004) Hemispheric roles in the perception of speech prosody. Neuroimage 23:344-357.

Hämäläinen M, Hari R, lmoniemi RJ, Knuutila JE, Lounasmaa OV (1993) Magnetoencephalography — theory, instrumentation, and applications to non invasive studies of the working human brain. Rev Modern Phys 65:413-497.

Hose B, Langner G, Scheich H (1983) Linear phoneme boundaries for German synthetic two-formant vowels. Hear Res 9:13-25.

Jusczyk PW (2002) Some critical developments in acquiring native language sound organization during the first year. Ann Otol Rhinol Laryngol Suppl 189:11-15.

Kohonen T, Hari R (1999) Where the abstract feature maps of the brain might come from. Trends Neurosci 22:135-139.

Koyama S, Akahane-Yamada R, Gunji A, Kubo R, Roberts TP, Yabe H, Kakigi R (2003) Cortical evidence of the perceptual backward masking effect on /l/ and /r/ sounds from a following vowel in Japanese speakers. Neuroimage 18:962-974.

Krumbholz K, Patterson RD, Seither-Preisler A, Lammertmann C, Lütkenhöner B (2003) Neuromagnetic evidence for a pitch-processing center in Heschl's gyrus. Cereb Cortex 13:765-772.

Liebenthal E, Binder JR, Spitzer SM, Possing ET, Medler DA (2005) Neural substrates of phonemic Perception. Cereb Cortex 15:1621-1631.

Lounasmaa OV, Hämäläinen M, Hari R, Salmelin R (1996) Information processing in the human brain:magnetoencephalographic approach. Proc Natl Acad Sci USA 93:8809-8815.

Lütkenhöner B, Krumbholz K, Lammertmann C, Seither-Preisler A, Steinstrater O, Patterson RD (2003) Localization of primary auditory cortex in humans by magnetoencephalography. Neuroimage 18:58-66.

Mäkelä AM, Alku P, Tiitinen H (2003) The auditory N1m reveals the left-hemispheric representation of vowel identity in humans. Neurosci Lett 353:111-114.

Maye J, Werker JF, Gerken L (2002) Infant sensitivity to distributional information can affect phonetic discrimination. Cognition 82: B101-B111.

Näätänen R, Picton T (1987) The N1 wave of the human electric and magnetic response to sound:a review and an analysis of the component structure. Psychophysiology 24:375-425.

Näätänen R, Winkler I (1999) The concept of auditory stimulus representation in cognitive neuroscience. Psychol Bull 125:826-859.

Narain C, Scott SK, Wise RJ, Rosen S, Leff A, Iversen SD, Matthews PM (2003) Defining a left-lateralized response specific to intelligible speech using fMRI. Cereb Cortex 13:1362-1368.

Obleser J, Elbert T, Lahiri A, Eulitz C (2003a) Cortical representation of vowels reflects acoustic dissimilarity determined by formant frequencies. Brain Res Cogn Brain Res 15:207-213.

Obleser J, Lahiri A, Eulitz C (2003b) Auditory-evoked magnetic field codes place of articulation in timing and topography around 100 milliseconds post syllable onset. Neuroimage 20:1839-1847.

Obleser J, Elbert T, Eulitz C (2004a) Attentional influences on functional mapping of speech sounds in human auditory cortex. BMC Neurosci 5:24.

Obleser J, Lahiri A, Eulitz C (2004b) Magnetic brain response mirrors extraction of phonological features from spoken vowels. J Cogn Neurosci 16:31-39.

Obleser J, Boecker H, Drzezga A, Haslinger B, Hennenlotter A, Roettinger M, Eulitz C, Rauschecker JP (2005) Vowel sound extraction in anterior superior temporal cortex. Hum Brain Mapp 10.1002/hbm.20201.

Oldfield RC (1971) The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9:97-113.

Pantev C, Bertrand O, Eulitz C, Verkindt C, Hampson S, Schuierer G, Elbert T (1995) Specific tonotopic organizations of different areas of the human auditory cortex revealed by simultaneous magnetic and electric recordings. Electroencephalogr Clin Neurophysiol 94:26-40.

Pena M, Bonatti LL, Nespor M, Mehler J (2002) Signal-driven computations in speech processing. Science 298:604-607.

Peterson G, Barney H (1952) Control methods used in a study of the vowels. J Acoust Soc Am 24:175-184.

Rauschecker JP, Tian B (2000) Mechanisms and streams for processing of 'what' and 'where' in auditory cortex. Proc Natl Acad Sci USA 97:11800-11806.

Roberts TP, Gage N (2002) M100 Latency tracks perception through a continuum of vowels. In:Proceedings of the 13th International Conference on Biomagnetism (Nowak H, Haueisen J, Gießler F, Huonker R, eds), p. 52. Berlin: VDE.

Robson MD, Dorosz JL, Gore JC (1998) Measurements of the temporal fMRI response of the human auditory cortex to trains of tones. Neuroimage 7:185-198.

Sanders LD, Newport EL, Neville HJ (2002) Segmenting nonsense: an event-related potential index of perceived onsets in continuous speech. Nat Neurosci 5:700-703.

Sarvas J (1987) Basic mathematical and electromagnetic concepts of the biomagnetic inverse problem. Phys Med Biol 32:11-22.

Scott SK, Johnsrude IS (2003) The neuroanatomical and functional organization of speech perception. Trends Neurosci 26:100-107.

Scott SK, Blank CC, Rosen S, Wise RJ (2000) Identification of a pathway for intelligible speech in the left temporal lobe. Brain 123:2400-2406.

Shestakova A, Brattico E, Soloviev A, Klucharev V, Huotilainen M (2004) Orderly cortical representation of vowel categories presented by multiple exemplars. Brain Res Cogn Brain Res 21:342-350.

Tanaka H, Fujita N, Watanabe Y, Hirabuki N, Takanashi M, Oshiro Y, Nakamura H (2000) Effects of stimulus rate on the auditory cortex using fMRI with 'sparse' temporal sampling. Neuroreport 11:2045-2049.

Wolfinger RD (1997) An example of using mixed models and PROC MIXED for longitudinal data. J Biopharm Stat 7:481-500.

Zatorre RJ, Belin P, Penhune VB (2002) Structure and function of auditory cortex: music and speech. Trends Cogn Sci 6:37-46.

Zielinski BA, Rauschecker JP (2000) Phoneme-specific functional maps in the human superior temporal cortex. Soc Neurosci Abstr 26:1969.