Specialization of left auditory cortex for speech perception in Man depends on temporal coding
Catherine Liegeois-Chauvel, Jozina B. de Graaf, Virginie Laguitton, Patrick Chauvel

To cite this version:
Catherine Liegeois-Chauvel, Jozina B. de Graaf, Virginie Laguitton, Patrick Chauvel. Specialization of left auditory cortex for speech perception in Man depends on temporal coding. Cerebral Cortex, Oxford University Press (OUP), 1999, 9, pp.484-496. hal-01454129

HAL Id: hal-01454129
https://hal.archives-ouvertes.fr/hal-01454129
Submitted on 22 Nov 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Specialization of Left Auditory Cortex for Speech Perception in Man Depends on Temporal Coding

Speech perception requires cortical mechanisms capable of analysing and encoding successive spectral (frequency) changes in the acoustic signal. To study temporal speech processing in the human auditory cortex, we recorded intracerebral evoked potentials to syllables in right and left human auditory cortices including Heschl's gyrus (HG), planum temporale (PT) and the posterior part of superior temporal gyrus (area 22). Natural voiced (/ba/, /da/, /ga/) and voiceless (/pa/, /ta/, /ka/) syllables, spoken by a native French speaker, were used to study the processing of a specific temporally based acoustico-phonetic feature, the voice onset time (VOT). This acoustic feature is present in nearly all languages, and it is the VOT that provides the basis for the perceptual distinction between voiced and voiceless consonants. The present research tests the hypothesis that left hemisphere's contribution to speech perception is based on the representation of formant transitions, rather than a deficit in the linguistic nature of formant transitions. This hypothesis could explain phonological disorders observed in language-learning-impaired children (Schwartz and Tallal, 1980), this temporal processing is likely to be a fundamental mechanism in sound perception which could be supported by cortical processing involving the auditory cortex.

One of several temporal cues used in perception and discrimination of stop consonants is the voice onset time (VOT). Lisker and Abramson (Lisker and Abramson, 1964) have shown that the voicing and aspiration differences among stop consonants in a wide variety of languages can be characterized by changes in VOT, which, in turn, reflect differences in the timing of glottal closure activity relative to supralaringeal events. It has been assumed that the perception of the VOT is under the control of the left hemisphere (Liberman et al., 1952; Lane, 1965, Fujisaki and Kawashima, 1971; Studdert-Kennedy, 1976; Ades, 1977; Miller, 1977; Pastore et al., 1977; Stevens, 1981; Kuhl and Padden, 1983; Macmillan, 1987). Several studies, however, have reported slight deficits in VOT discrimination in patients with damage to the left hemisphere (Oscar-Berman et al., 1975; Basso et al., 1977; Blumstein et al., 1977; Miceli et al., 1978; Itoh et al., 1986). Electrophysiological data suggest that the perception of the VOT is controlled by several cortical processes — some of which are restricted to the right hemisphere and others of which are common to both hemispheres (Molfese, 1978). A question which we would like to address is whether the VOT may be under the control of the left hemisphere (Oscar-Berman et al., 1975; Basso et al., 1977; Blumstein et al., 1977; Miceli et al., 1978; Itoh et al., 1986).

Introduction

Language perception is essentially based on subtle differences in the timing of acoustic elements of the speech signal. Many neurological observations in aphasic patients have shown that disorders in auditory language perception are linked to an impaired processing of time-related (temporal) information (Efron, 1963; Robin et al., 1990). It has been postulated that an enhanced capacity to analyse and encode successive temporal changes in the acoustic signal, requiring cortical mechanisms, may underlie the left hemisphere's contribution to speech processing (Tallal and Piery, 1973; Tallal and Newcombe, 1978). These authors demonstrated that a deficit in the processing of the rate of change of acoustic cues, such as the formant transitions, rather than a deficit in the linguistic nature of the stimuli, could explain phonological disorders observed in language-learning-impaired children. This hypothesis seems to be confirmed by the improvement of speech discrimination and language understanding in these children after a period of intensive behavioral training consisting of recognition of brief and fast sequences of speech and non-speech stimuli (Merzenich et al., 1996; Tallal et al., 1996). Since disturbances in temporal processing can create deficits in language-learning-impaired children (Schwartz and Tallal, 1980), this temporal processing is likely to be a fundamental mechanism in sound perception which could be supported by cortical processing involving the auditory cortex.

In this study, we examine the neural responses to speech and
non speech-sounds within the cortical auditory areas such as HG, PT and the anterior region to HG corresponding to area 22 (Brodmann, 1909). We focus on the stop consonants to investigate the nature of the temporal mechanisms involved in consonant perception. In particular, we seek to assess the degree to which the acoustic versus phonetic nature of the VOT is coded by the different auditory areas, leading to cortical lateralization in the perception of stop consonants.

Materials and Methods

Subjects
Seventeen epileptic patients (eight male, nine female, aged 17–38 years) participated to this study. It is important to note that these patients have been a posteriori selected in such a way that none of them had their epileptogenic zone including the auditory areas, and no patient showed atypical language representation during specific examinations. Recordings of brainstem evoked potentials carried out before the stereoelectroencephalography (SEEG) confirmed that all the patients had normal afferent pathway conduction.

All patients were informed about the research protocol during the SEEG and gave consent (Ethical Committee, Rennes, November 15, 1993).

Presurgical Exploration
The presurgical SEEG (Bancaud et al., 1965; 1992; Chauvel et al., 1996) of patients with temporal, temporo-parietal or temporo-frontal epilepsy requires implantation of multiple depth electrodes in the temporal lobe and neighbouring areas of adjacent lobes, its purpose is to determine the anatomical structures involved in the initiation and propagation of seizures and the accurate limits of future cortical excision (Talairach et al., 1974). In addition to the cases where the posterior areas of the superior temporal gyrus (STG) are suspected of constituting the whole or part of the epileptogenic zone, this region is frequently explored because of its strategic position as a pathway for ictal discharge propagation towards inferior parietal and opercular regions (Bancaud and Chauvel, 1987), as well as for the necessity for functional mapping before surgery.

Anatomical Definition of Depth Electrode Position
The multilead electrodes (0.8 mm diameter, 5–15 contacts, 2 mm length, 1.5 mm apart) are introduced orthogonally through the usual double grid system fastened to the Talairach stereotaxic frame. Anatomical localization of each lead of the electrode is based on a stereotaxic method which has been described elsewhere (Szikla et al., 1977; Talairach and Tournoux, 1988; Liégeois-Chauvel et al., 1991). Owing to the oblique orientation of HG, a single electrode can explore different auditory areas (medial and/or lateral part of HG, and/or PT and/or area 22; see Table 1).

We could verify that the selected anatomical structures had indeed been explored by visualization of the electrode tracks through stereotactic magnetic resonance imaging (MRI) (Fig. 1).

Figure 2 shows a typical brain (from an MRI) in a sagittal (Fig. 2A, B) and frontal view (Fig. 2C). Figure 2A, B shows the entrance points of the electrodes used to explore the left and right STG; the numbers on the points correspond to the cases listed in Table 1. Figure 2C shows the points at which the leads of the electrodes were most responsive, in selected cases, during the exploration of the posterior portion of the STG.

Stimuli
The auditory stimuli used were speech sounds and speech analogue sounds.

Speech Sounds
We used stop consonants (/ba/, /da/, /ga/, /pa/, /ta/, /ka/) pronounced and recorded by a native French speaker in a sound-attenuated room. It is
possible to describe stop voiced consonants in terms of a sequence of acoustic segments. These segments appear on the spectrogram as a series of relatively homogeneous stretches of the signal demarcated by rather abrupt changes in spectral form, marking the boundary between segments (Fant, 1973). In French, the VOT (i.e. the temporal relationship between the release burst in a stop consonant and onset of glottal pulsing) of the voiced consonants (/b/, /d/, /g/) is negative, preceding by ∼110 ms the release burst, whereas for voiceless consonants (/p/, /t/, /k/), the VOT is positive, following the release burst (20 ms) (left side of Fig. 3).

Note that there are timing differences in production and perception of VOT across different languages; for English voiced stops, voicing precedes or is nearly coincidental with release (Lisker and Abramson, 1964).

The three voiced syllables (VS) (/ba/, /da/, /ga/) had a VOT of ∼110 ms and a total length of ∼360 ms. The three voiceless syllables (VLS) (/pa/, /ta/, /ka/) had a length of ∼200 ms. The syllables were presented to all the patients over headphones in a pseudo-random order. The stimulus intensity was adjusted to 70 dB SL (sensation level, relative to the threshold at 1000 Hz).

Acoustic Speech Analogues
The four acoustic speech analogues used are displayed at the right side of Figure 3. VOTa, the sound mimicking the voicing, lasts 110 ms and contains three low frequencies (200, 400, 600 Hz). VLa-short, the complex-sound analogue to a vowel, reflects the vowel’s spectral content (200, 400, 600, 800, 1500, 2700, 4100 Hz) and duration (225 ms). Va mimics the syllable’s temporal course: it has 110 ms of near silence followed by seven stronger frequencies present in the two previous complex sounds for a total duration of 320 ms. VLa-long has the same spectral components as Va but without the near silence (pre-voicing) at the beginning; its total duration is 400 ms.

All the patients had heard VOTa and VLa-short sounds and either Va or VLa-long sounds, delivered in a pseudo-random order in the same series.

Recordings
We recorded from 65 different sites of recording in auditory areas: 14 and 23 leads in the right and left HG respectively, 4 and 7 leads in the right and left PT, and 9 and 8 leads in right and left area 22, located anteriorly to HG.

The recordings of intracerebral AEPs were monopolar, with each lead
of each depth electrode referenced to an extra-dural lead. All the signals were amplified and filtered simultaneously (bandpass 1 Hz–1.5 kHz). Data acquisition started 82 ms before the presentation of the sound and lasted for 819 ms. Average waveforms (100 trials) were digitally low-pass filtered using a Butterworth simulation (cutoff frequency 175 Hz, slope 48 dB/octave).

During a recording session, each patient laid comfortably in a chair in a sound-attenuated room and listened passively to the sounds.

Results

Recordings in the Left HG
No difference was found in AEPs among the three VS nor among the three VLS. Therefore, we present the grand average of AEPs in response to voiced versus voiceless stimuli at different recording sites.

Figure 4A–H illustrates the pattern of AEPs recorded from eight selected leads at various sites in the left HG to the VS (solid curves) and VLS (dashed curves). The basic sequence of cortical activation in response to both sets of stimuli was similar up to 170 ms mean latency. However, two main features of the intracortical responses are modified by VOT. First, the response seemed to be more complex for VS than for VLS. Except for the earliest components recorded exclusively in the primary cortex, six main successive components (48, 93, 168, 207, 262 and 353 ms mean latency) are recorded in response to VS and only five to VLS (45, 88, 163, 216, 292 ms).

Second, the processing of the VS in the left HG is sequential in nature. There are prominent evoked responses (N/P48, N/P93) at the onset of the voiced consonants. The next response (N/P170) is time-locked with the release burst of the consonant, followed by the responses N/P207 and N/P262, which correspond to the processing of the last part of the syllable. The last component, a slow-wave positive component peaking around 353 ms, marks the end of the response. Note that the interval separating the early responses and the N/P170 equals the duration of VOT.

Occasionally, the offset of the sound evoked a response that peaked at 277 ms for VLS and at 420 ms for VS. Interestingly, multiple low-voltage components superimposed over the main components were seen often for VS and occasionally for VLS. These oscillatory components appeared during the formant transitions after the burst and lasted almost the duration of the vowel (Fig. 4A,F). The power spectra of these responses peak at a frequency of ∼80 Hz.

Recordings in the Right HG
In the right HG (Fig. 5), VS and VLS elicited response patterns that were similar in the order and duration of response components. Components always present for both sets of stimuli were at 46, 84, 173 and 283 ms (mean latency). The shape of responses depended on neither the duration nor the nature of the stimulus. Thus, the sequential processing of the different
parts of VS observed in the left HG is not found in the right HG. Moreover, there are no oscillatory responses.

Recordings in PT
We analyzed the evoked responses to VS and VLS in the left and
right PT and area 22 in order to see if the asymmetry of processing is preserved in these regions.

In the left PT (Fig. 6A,B), evoked responses to VS (full curves) were similar to those recorded in HG, and the same successive components were observed (70, 96, 161, 221, 252, 337 ms mean latency). The first component peaked later than the first one found in HG. The temporal processing of VOT and the vowel seemed to be preserved. However, the difference in the morphology of evoked responses between VS and VLS was less clear. In some cases (Fig. 6A), a large slow wave culminating between 500 and 600 ms after stimulus onset could be recorded.

Figure 6C shows the responses recorded in the right PT. No difference is observed in waveforms of evoked responses to both stimuli.

Recordings in Area 22

The results of recordings in left and right area 22 are displayed in Figure 7A–F. For both VS and VLS, responses had a very similar waveform, a triphasic potential in which the latencies of different components varied as a function of the precise position of the electrode. Any sequential processing did not emerge from these data, as similar AEPs were recorded for both types of syllables.

Tables 2 and 3 sum up data obtained from recordings in all the patients. For clinical considerations and owing to individual anatomical variations, the number of leads and the placement of the electrodes differed across patients. Therefore, a grand average of responses recorded at different leads, as is often done with scalp recordings, would be of little significance. Nonetheless, we have labelled the peak latencies of the main components evoked by both types of syllables on the 65 responsive leads in order to check the validity of this temporal coding. The mean latencies and SDs of peaks are summarized in Table 2 for left and in Table 3 for right auditory areas. The reproducibility of the AEPs is manifested in the relatively small SDs, which varied from 5% for the short latencies to 20% for the long latencies.

As we could observe in selected recordings displayed in

Figure 5. AEPs recorded in the right HG. See the legend of Figure 4 and comments in the text.
**Figure 6.** AEPs recorded at different sites within left and right PT in response to VS (full curve) and VLS (dashed curve). See legend of Figure 4 and comments in the text.

**Table II**
Mean latency (± SD) of the successive positive and negative main components evoked by voiced (normal type) and voiceless (bold type) syllables recorded in different anatomical structures in the left auditory cortex

<table>
<thead>
<tr>
<th>Structure</th>
<th>Primary components</th>
<th>Last component</th>
</tr>
</thead>
<tbody>
<tr>
<td>HGm</td>
<td>25.4 ± 2.7</td>
<td>166 ± 6.3</td>
</tr>
<tr>
<td>HGl</td>
<td>51.7 ± 8.7</td>
<td>170 ± 19.3</td>
</tr>
<tr>
<td>PT</td>
<td>47.3 ± 6.7</td>
<td>168 ± 2.9</td>
</tr>
<tr>
<td>Area 22</td>
<td>65.6 ± 5.4</td>
<td>176 ± 7.9</td>
</tr>
</tbody>
</table>

*Indicates the supplementary negative component found for the voiced syllables in HG and PT (see Figs 4 and 6). Corresponding to earlier results (Liegeois-Chauvel et al., 1991), the primary components were only recorded in the medial part of HG. Notice the difference in latency of the last component recorded in HG between voiced and voiceless syllables.

**Table III**
Mean latency (± SD) of the successive positive and negative main components evoked by voiced (normal type) and voiceless (bold type) syllables recorded in different anatomical structures in the right auditory cortex: contrary to left HG, only a very small difference in latency of the last component between voiced and voiceless syllables was found for right HG.

<table>
<thead>
<tr>
<th>Structure</th>
<th>Primary components</th>
<th>Last component</th>
</tr>
</thead>
<tbody>
<tr>
<td>HGm</td>
<td>23.3 ± 1.5</td>
<td>97 ± 12.8</td>
</tr>
<tr>
<td>HGl</td>
<td>47.3 ± 2.3</td>
<td>101 ± 2.9</td>
</tr>
<tr>
<td>PT</td>
<td>46.3 ± 7.9</td>
<td>165 ± 5</td>
</tr>
<tr>
<td>Area 22</td>
<td>60 ± 3.6</td>
<td>164 ± 6.7</td>
</tr>
</tbody>
</table>

*Indicates the supplementary negative component found for the voiced syllables in HG and PT (see Figs 4 and 6). Corresponding to earlier results (Liegeois-Chauvel et al., 1991), the primary components were only recorded in the medial part of HG. Notice the difference in latency of the last component recorded in HG between voiced and voiceless syllables.
Figures 4 and 6, the processing of voiced and voiceless syllables is different in the left HG and PT. This is confirmed by the omnipresence of the negative component peaking at 252–262 ms preceded by positivity (204–221 ms) for VS. The last component latency is longer for VS than for VLS, suggesting that the syllable duration is involved in the temporal processing. This processing did not take place in the right auditory areas, so AEP components in response to VS or VLS were quite similar.

The present results demonstrate that the temporal coding of speech sound is performed in the left HG and PT and not in the right homologous cortical structures. This striking event highlights the functional difference between the area 22, on the one hand, and HG and PT, on the other, since this temporal coding is not observed in the left area 22.

**Evoked Responses to Acoustic Speech Analogues**

To investigate whether this temporal processing is related to the acoustic or to the linguistic features of the syllable, we analysed evoked responses to acoustic speech analogues of different durations and spectral composition (see Materials and Methods).

Figure 8 displays evoked responses to complex tones within the different auditory cortical areas. The first phenomenon is the waveform similarities of potentials from the right side whatever the duration and the spectral complexity of sound. In contrast, evoked responses from the left side were modulated by the stimulus duration. A striking example was seen in Figure 8A, E. An off-response occurred in all three recordings in the left HG (Fig. 8A) and in none of the recordings in the right HG (Fig. 8E). The off-response was correlated with the duration of the sound, beginning at 400, 270 and 180 ms for VLa-long (solid curve), VLa-short (dashed curve) and VOTa (dotted curve) respectively.

The solid curves in Figure 8B–D show that Va evoked a double response in the left HG (just as did the natural VS): an initial response to sound onset and a second to burst onset. The
response to the burst, however, was different for speech and non-speech sounds; the oscillatory responses superimposed on the potentials evoked by a syllable were never evoked by an acoustic speech analogue.

Sequential processing of the complex sound was predominantly observed in the left HG, to a lesser degree in the left PT, and not in area 22.

Discussion
The results of this study suggest that, in humans, evoked responses of left auditory cortex are related to the temporal structure of sounds. Indeed, we found that voiced and voiceless consonants were differently processed in left HG, suggesting a specific temporal processing of sounds which was not present in the right homotopic structures. It appears that in left HG,
neurons responsible for representing the different acoustic elements in syllables were activated in succession. In contrast, in the right cortical auditory areas, speech sounds evoked a similar pattern of response whatever the nature of the syllable, as if the syllable was processed as an entity.

A differential coding of voiced and voiceless syllables is preserved in the left PT. In contrast, this specificity has not been found for area 22, located anteriorly to HG, which does not show any asymmetry between both hemispheres in the processing of syllables. These results raise the question of the functional role of the different cortical auditory areas.

Time-locked responses are evoked to the consonant onset as well as to the release burst of the French VS in the left HG. Our data support a temporally based physiological mechanism for the differential perception of VS and VLS and agree with suggestions that this distinction is partially based upon detection of the non-simultaneous onset of the VOT and the release burst (Stevens, 1980). Furthermore, Dwyer et al. (Dwyer et al., 1982) have shown that the presence of abrupt onset for stop consonants crucially determined lateralized processing. This mechanism is specifically involved in the left auditory cortex. Similar time-locked responses to speech sounds are also seen in scalp recordings (Kaukoranta et al., 1987). But these AEPs were bilaterally distributed. This absence of asymmetry could be explained by the fact that HG is deeply buried inside the sylvian fissure and that the specific activity of primary area, localized in the dorso-postero-medial part of HG (Liegosi-Chauvel et al., 1991) is difficult to record from surface-recording techniques (EEG as well as MEG). At the level of the scalp surface, most of the recordable evoked responses presumably originate in cortical regions near the recording electrodes. However, relatively large signals originating from PT and area 22 might often make a significant contribution to the activity observed at a given scalp recording site.

This processing seems to be related to the acoustic rather than to the phonetic composition of the sound since identical time-locked responses were recorded to speech sound and non-speech sound with the same time course. Indeed, similar time-locked responses as those for speech are seen for the speech analogue sounds, implying that these responses are not specific detectors for acoustic speech components. This had already been observed in Kaukoranta’s study (Kaukoranta, 1987). Right ear advantage (REA) has been found for auditory sequences suggesting left-hemisphere specialization for processing stimuli characterized by temporal acoustic features (Divenyi and Efron, 1979; Leek and Brandt, 1983). Schwartz and Tallal (Schwartz and Tallal, 1980) showed a significantly greater REA for discriminating stop consonants that contain transient transitions within the speech stimuli, that were not incorporated into the non-speech analogue sounds and not for the speech analogue sounds could be interpreted in two ways. First, it is possible that the formantic transitions within the speech stimuli, that were not incorporated into the non-speech analogue stimuli, might have driven the lateralized oscillatory responses. Eggermont (Eggermont, 1994) showed that the neural synchronization coefficient increased more during stimulus presentation as compared with spontaneous activity, especially when the stimulus was an amplitude- and/or frequency-modulated burst. Second, since the high-frequency superimposed responses are observed only in response to speech sounds and recorded in the left auditory cortex, this oscillatory activity could represent binding functions among the different acoustic features, each of which is detected by different individual neurons leading to a unified meaningful percept. Similar oscillations have been extensively studied since the work of Galambos et al. (Galambos et al., 1981) as gamma-band activity supporting general perception mechanisms in the central auditory system (Mäkelä and Hari, 1987; Panetz, 1995). However, DeCharms and Merzenich (DeCharms and Merzenich, 1996) showed that neurons in the primary auditory cortex can coordinate the relative timing of their action potentials during the entire duration of continuous stimuli. They demonstrated that the relative timing of cortical action poten-
tials can signal stimulus acoustic features themselves, a function even more basic that perceptual feature grouping.

The present work has yielded direct evidence that high-frequency responses are generated in the human auditory cortex [confirming our previous data (Liégeois-Chauvel et al., 1994)]. Future studies using speech analogues with other acoustic features may shed light on our understanding of the mechanisms underlying the high-frequency responses.

Relevance to Human Anatomy

This functional asymmetry between the left and right HG may be related to anatomical differences. It is well known that the morphological asymmetry observed in the posterior temporo-sylvian region in man [especially in the associative auditory areas, such as the PT (Braak, 1978; Galaburda and Sanides, 1980; Steinmetz et al., 1989; Rademacher et al., 1993; Kulychny et al., 1994)] is related to the left hemisphere specialization for speech. Seldon (Seldon, 1981) found that, although the total area of primary cortex is the same on both sides, neurons in the left HG generally have a larger tangential dendritic arborization extent in the cell columnar organization than those in the right HG. Recently, morphological imagery studies showed an anatomical asymmetry that arises from a difference in the volume of white matter, which is larger in left than in right HG, which could be due to the presence of thicker or a larger number of myelinated fibers (Penhune et al., 1996). The larger volume of cortical white matter of left HG, involving a greater number of connecting fibres and/or more myelinated fibres, could be responsible for the observed enhanced temporal processing capability in left HG.

Impairment in Temporal Processing Observed in Language Disorders

Left-hemisphere or bilateral lesions to auditory cortices lead to word deafness. Not all aspects of speech discrimination are disrupted to the same extent in these pathological situations. There is a general agreement that the discrimination of steady-state vowels is preserved, while discrimination of consonants, especially stops, is not (Saffran et al., 1976; Auerbach et al., 1982; Mendez and Gheethan, 1988; Yaqub et al., 1988). Recently, evidence has been provided suggesting that pure word deafness may be a manifestation of a basic temporal processing disorder. Two studies have questioned the ability of word-deaf patients to discriminate VOT. Auerbach et al. (Auerbach et al., 1982) reported a patient with bilateral lesions who showed a normal labelling function but an impaired discrimination of consonant–vowel syllables when VOT differed by only 20 ms, even when the syllables had VOT values close to those defining the phonetic boundary between the relevant voiced and voiceless stop consonants. This suggests an underlying deficit in the temporal processing of successive acoustic events. Miceli (Miceli, 1982) presented comparable data for a patient with a more generalized cortical auditory disorder whose deficit in the discrimination of verbal and non-verbal sounds could be at least partly explained by confusions in the temporal processing of successive acoustic events [for review see Phillips and Farmer (Phillips and Farmer, 1990)]. These clinical data are in agreement with our results demonstrating that a non-verbal sound is also temporally coded in the left auditory cortex.

As already mentioned above, since speech signals have a complex time course, a strong performance in auditory temporal processing is necessary in order to process speech information. The level of temporal representation used by the nervous system in the processing of speech signals is an open question. Indeed, the discrimination of steady-state vowels is maintained after bilateral lesions of the primary auditory cortex (Auerbach et al., 1982; Von Stockert, 1982; Mendez et al., 1988). The waveforms of these phonemes are quite different. Vowels have no time structure above the level of the periodicities in their waveforms, and the fact that they are perceived after bilateral lesions presumably reflects the activation of other cortical auditory fields unaffected by the pathology. In contrast, stop consonants contain many transient acoustic events that are closely spaced in time (range around tens milliseconds). Perhaps, the currently most valid theory of speech recognition is the acoustic one: it is based on the fact that each phoneme has a unique acoustic signature (short-term spectrum and temporal variations) (Blumstein and Stevens, 1979, 1980) and that the perceptual mechanisms use a sliding temporal window to sample the stream of sound. The resulting samples then are matched against internal templates for phonetic tagging (Moore et al., 1988; Viemeister and Wakefield, 1991; Phillips, 1993b).

In conclusion, this enhanced sensitivity to temporal acoustic characteristics of sounds of left human auditory cortex probably reflects information processing necessary for further speech processing. The task of the left primary and secondary auditory cortex would be to provide the cerebral language processor with a representation of the acoustic signal that has sufficient spectral and temporal pertinence to permit the phonetic tagging. That could explain why the perceptual elaboration of sounds with the relevant time frame of auditory content should be so dependent on the auditory cortex as is shown in language disorders.

A remaining question is why the auditory temporal processing in man should be more developed in the left auditory cortex whereas this mechanism is bilateral in monkey. Given the fact that the processing of syllables in the left hemisphere is already present in 3 month old infants (Dehaene-Lambertz and Dehaene, 1990), further research is needed to reveal to what extent the innate brain mechanisms of speech perception can be modulated by learning. This will give important knowledge about cortical plasticity.

Notes

We would like to thank Professors O. Sabouraud and P. Buser for critically commenting upon a previous version of the paper, Professor J.M. Scarabin for performing the MRIs, Dr J. Regis for illustration (three-dimensional rendering of the brain), P. Marquis for help in experiments and Dr B. Scharf for English revision. Address correspondence to C. Liégeois-Chauvel, Laboratoire de Neurophysiologie et Neuropsychologie, INSERM CJF 9706, Université de la Méditerranée, 27 Bd Jean Moulin, 13385 Marseille Cedex 5, France. Email: liegeois@arles.timone.univ-mrs.fr.

References


