

PRE-ATTENTIVE PROCESSING OF CONSONANT-VOWEL SYLLABLES AT THE SUPRATEMPORAL PLANE: A WHOLE-HEAD MEG STUDY

Klaus Mathiak, Ingo Hertrich, Werner Lutzenberger, and Hermann Ackermann
University of Tübingen, Germany

ABSTRACT

The present study investigated brain hemisphere lateralization effects with respect to the encoding of dynamic aspects of the acoustic speech signal. Using a whole-head device, auditory evoked magnetic fields were recorded in response to vowels as well as syllable-like structures differing either in direction or in duration of the formant transients. All trials used a synthesized vowel as a template against which the deviants had to be matched. Both the N1m-component and difference waves between the magnetic fields to standard and respective rare events (MMNm) were calculated. (a) Vowel mismatch (/a/ against /e/) resulted in enlarged N1m-amplitude reflecting, most presumably, peripheral adaptation processes. (b) As concerns syllable-like structures differing in their dynamic characteristics, only the shortest transient duration (= 10 ms) elicited a significant lateralization effect toward the left hemisphere. Thus, pre-attentive acoustic processing already yields lateralization effects and does not seem to depend on encoding of linguistic categories.

1. INTRODUCTION

Virtually all well-examined split-brain patients exhibited to some degree auditory language comprehension by the right hemisphere [1]. Rather than exclusively mediating speech perception, the dominant perisylvian cortex just seems to be more proficient in this regard. Especially, the encoding of consonantal sounds has been found to predominantly depend upon the left side of the brain. Using dichotic pairs of consonant-vowel-consonant (CVC) syllables differing in a single phone each, e.g., /tap/ and /kap/, Studdert-Kennedy and Shankweiler [2] observed a significant right-ear advantage (REA), i.e., left-hemisphere superiority, for accurate recognition both of the initial and final sound. Clinical data corroborate the notion of different speech sound processing mechanisms.

The various stop consonants (plosives) of the English or German language, i.e., /b/, /p/, /d/, /t/, /g/, and /k/, are cued by shifts of spectral energy distribution (transients) extending across a few tens of milliseconds [3]. These dynamic features of spectral energy distribution signal place of articulation of the plosives. On the basis both of clinical and experimental data, Tallal, Miller, and Fitch [4] proposed left-side superiority of consonant processing to reflect a higher proficiency of the dominant hemisphere in decoding formant transients. For example, selective damage to the left, but not the right cerebral hemisphere compromises the discrimination between speech sounds incorporating rapidly changing spectral cues [5]. Furthermore, dichotic presentation of syllables with transients of a rather short duration (= 40 ms) yielded a significant REA in normal subjects [6]. Stimuli comprising extended transitional elements (= 80 ms) failed this effect.

Whereas these findings referred to support the notion of a left-hemisphere specialization for the analysis of the dynamic aspects of the speech signal, time course and intra-

hemispheric location of the underlying mechanism remain to be established. Animal experimentation found precise neural encoding of the acoustic correlates of speech sounds such as voice onset time or formant transients within the auditory cortex [7]. For example, the syllables /ba/ and /da/ elicit multiple unit activities, time-locked to the formant transients, at different sites of the involved areas. Assuming acoustic rather than linguistic processing mechanisms underlying superior left-hemisphere encoding of stop consonants as compared to vowels, these effects might be expected to arise already at the level of the supratemporal plane, i.e., during an early stage of cerebral speech sound encoding. This suggestion is further supported by data indicating a common supramodular device to mediate the analysis of dynamic auditory events both in the verbal and non-verbal domain [4]. At variance with this assumption, however, positron emission tomography (PET) studies found stimuli incorporating rapid spectral changes to elicit significant lateralization effects at the level of the frontal lobes rather than the temporal cortex [8]. As compared to PET, magnetoencephalography (MEG) might represent a more appropriate tool in order to test the hypothesis of early lateralized encoding of formant transients at the level of the supratemporal plane. First, MEG exhibits selective sensitivity to tangentially oriented neural currents and, therefore, predominantly focuses on activity of the acoustic cortex. Second, the temporal resolution of MEG is within the domain of milliseconds. The present study used the recently introduced technology of whole-head MEG which simultaneously obtains evoked magnetic fields from both hemispheres and therefore should allow to detect small but significant lateralization effects of cerebral functions.

Presentation of randomly interspersed rare auditory stimuli within a sequence of homogeneous events (oddball design) provides the basis for the computation of magnetoencephalographic mismatch negativity (MMNm), i.e., the difference wave of the magnetic fields elicited by the deviant and the frequent stimuli. MMNm has been shown to be sensitive to acoustic speech parameters. Dipole source analysis indicates MMNm to be generated at the level of the supratemporal plane [9].

Both experiments of the present study used a synthesized vowel of a male voice as the frequent event, providing an acoustic template against which the various deviants had to be matched. The first measurements comprised vowel /a/ as the standard and three types of deviant stimuli: the vowel /e/, the syllable /ba/, which incorporates an initial transient of the first three formants each followed by the same steady-state components as in case of the standard vowel, and the pseudo-syllable /ξa/ differing from /ba/ in the direction of the first formant transition (Figure 1). Since the human vocal tract is unable to produce the formant structure of /ξa/ (see [10] for similar stimuli), this stimulus exhibits an inherent ambiguity with respect to acoustic structure and linguistic

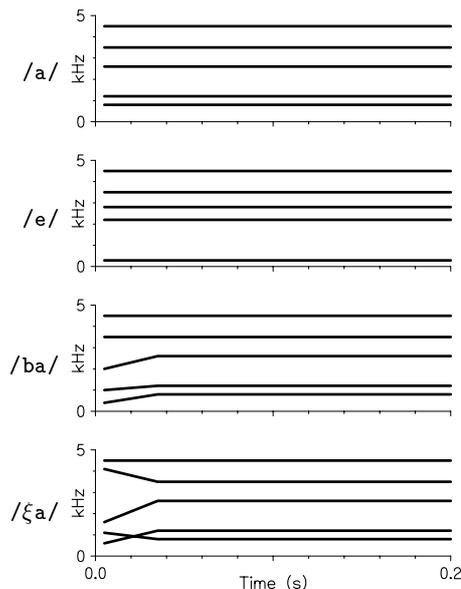


Figure 1: Stylized spectrograms of the synthesized stimuli: vowels /a/ and /e/ differ in the first two formants; stop consonants are characterized by initial transients. The formant transition of /ξa/ (crossing transients) cannot be produced by the vocal tract.

categorization. The second experiment systematically varied the duration of the second formant transition of the syllable /bi/ between 10 and 90 ms and tested detection of these events against vowel /i/ as a template (Figure 2).

On the basis of the model of Tallal and coworkers [4] on speech sound processing, it was hypothesized (a) that lateralization effects depend upon acoustic structure rather than perceived linguistic category, i.e., stimulus /ξa/ behaves like /ba/ and not like vowel /e/, and (b) that lateralization effects emerge only in the presence of short transitional elements.

Assuming superiority of the left perisylvian cortex for the analysis of formant transients, an enhanced matching procedure must be expected at the dominant side giving rise to lower threshold, shorter latency, or larger amplitude of the respective MMNm.

2. MATERIALS AND METHODS

2.1. Subjects

Altogether, 20 paid right-handed native speakers of standard German (age 26-40 years; median 31 years; 9 females) participated in the present study (experiment 1, $n=12$; experiment 2, $n=13$). At clinical examination, the volunteers showed unimpaired hearing sensitivity. None of them had a history of any relevant audiological or neurological disorder. Informed consent was obtained from all subjects.

2.2. Experimental procedure

The first experiment relied on the classical oddball design, i.e., the application of a randomized series of a frequent (80% of events) and a rare stimulus (20%), with the synthesized vowel /a/ as the standard in all instances. Three different deviants were tested in separate blocks each: vowel /e/,

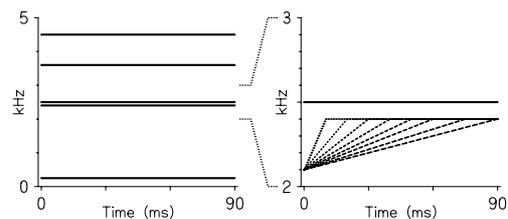


Figure 2: Left panel: stylized spectrogram of vowel /i/ (frequent event). Right panel: zoomed frequency range 2–4 kHz showing the various F2-transients (10–90 ms) as well as F3 of the deviant /bi/ stimuli. Note that cumulative spectral distance increases with transient length whereas speed of transition is decreasing.

syllable /ba/, and pseudo-syllable /ξa/. Figure 1 displays the spectral structure of these auditory events. Vowels /a/ and /e/ exhibit stationary spectral energy distribution across stimulus length but differ in the location of the first to fifth formants (F1–F5). In case both of /ba/ and /ξa/, an initial transient (duration = 35 ms) precedes the steady-state component of the lower formants. The fundamental and formant frequencies were selected, such that the synthesized vowel /a/ elicited the impression of a male voice. Vowel /e/ differed in F1 and F2.

The second experiment modified the classical oddball design: A set of eight deviant stimuli derived from the syllable /bi/ by systematic adjustment of the length of the F2-transient (transient durations 10–90 ms; see Figure 2) was randomly interspersed into a series of the standard vowel /i/. Altogether, the deviants amounted to 20% of the 450 events in each of the three blocks. As a rule, linguistic categorization of the synthesized stimuli varied depending upon transient length: The event with the shortest duration was perceived as a vowel concomitant with an initial click, intermediate transients sounded like syllable /bi/, the slowest change of spectral energy distribution yielded the impression of a diphthong.

2.3. Stimulus synthesis

In order to avoid eventual pitfalls of linear predictive coding (LPC), an additive synthesis algorithm was applied. Each of the five formants was modeled as an amplitude and frequency modulated sinusoid phase-locked to the fundamental frequency. This algorithm allows for continuous alteration of formant trajectories within single pitch periods and gives rise to homogeneous and intelligible vowel quality of the produced CV-syllables with defined spectral energy distribution.

2.4. Data acquisition and analysis

Auditory evoked magnetic fields were recorded by means of a 151-channel whole-head gradiometer. Subjects were instructed to ignore the stimuli. Experiment 1 and 2 slightly differed in sweep length (500 vs 548 ms), pre-stimulus baseline (48 vs 148 ms), onset-to-onset inter-stimulus interval (805 vs 605 ms), and stimulus length (200 vs 90 ms).

N1m-peak point was defined as the time-point of maximum global field power about 100 ms post-trigger. Source analysis was performed assuming a simple-sphere head model and two

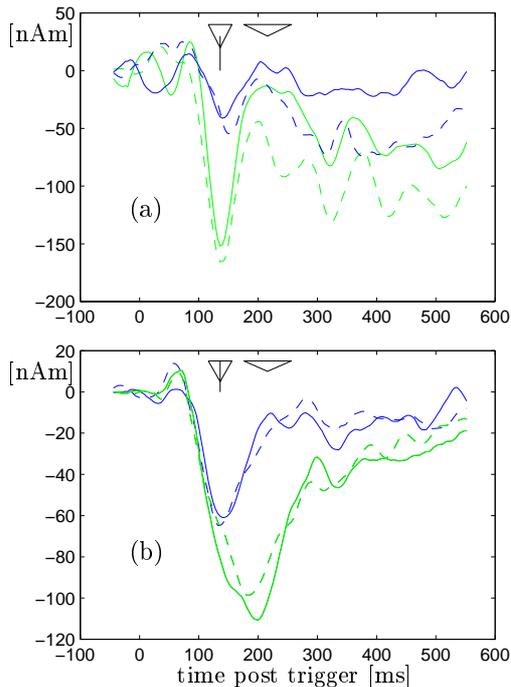


Figure 3: Time course of calculated dipole moments in a representative subject (straight line = left hemisphere, dashed line = right hemisphere). The N1m-peak is indicated by the vertical line in the small triangle; the broader triangle represents the MMNm template. Upper panel (a): averaged response for frequent /a/ (black; last stimulus prior to the deviant) and deviant /e/ (green). Note the large dipole moment in the rare condition at the time of the N1m-peak. Lower panel (b): average for frequent /a/ (blue) and deviant /ba/ (gray). Activation on these conditions diverge progressively after N1m-peak with maximum difference about 80 ms later.

dipole currents. The following sections consider the results of the fit representing locations of main electrical activity in the left versus right auditory cortex or the planum temporale. The subspace projection on the two dipole components provided a measure of cortical activity in terms of dipole strength at each time point. Based on the extracted data we calculated the averages for the distinct frequent and deviant stimuli separately. Quantification of MMNm and N1m-amplitude were done by weighting the mismatch curves with a triangular window (see Figure 3).

The measured MEG signal represents the sum of multiple signal and noise sources. Consequently, intra-subject variability was considered an additive error term. Since individual source configurations do not solely result in additive inter-subject variability due to diverse scaling properties, group statistics were derived from robust linear intra-subject measures in combination with the non-parametrical asymmetrical sign test. The linear intra-subject parameters included N1m-amplitude difference, MMNm, and the linear main effect of the transient durations, respectively. As a measure of relative lateralization effects, the interaction between hemisphere

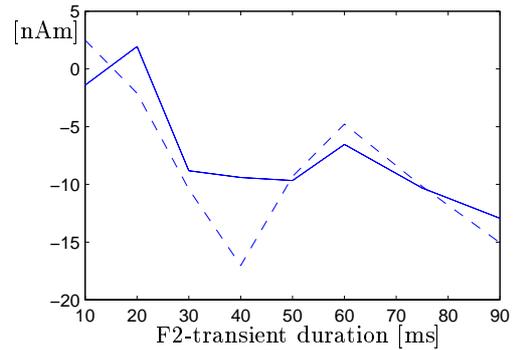


Figure 4: Average MMNm-amplitude across F2-transient duration of the deviant stimuli (straight line = left hemisphere, dashed line = right hemisphere). The amplitude of the mismatch reaction was determined 80 ms after N1m-peak. Negativation against baseline close to zero indicates stronger activation. The robust mean over the group illustrates the findings of non-parametrical statistics. MMNm increases with transient duration and shows additional bilateral enhancement within the domain of 30–40 ms as well as an unilateral left-sided response for transients of 10 ms duration.

and transient duration was considered.

3. RESULTS

3.1. Experiment 1

Subjects showed a global field power maximum within a time-interval extending from 96 to 125 ms post trigger (N1m-component). Bilateral dipoles originating at the supratemporal planes allowed to sufficiently explain the measured averaged magnetic fields. Statistical analysis of the N1m-components across group did not reveal any significant lateralization effects, neither with respect to amplitude nor latency (all $p \geq 0.2$). An enlarged N1m-amplitude of deviant vowel /e/ emerged ($p < 0.01$) when matched against the preceding frequent /a/ as a template (Figure 3a).

With vowel /a/ as the standard, the deviants /ba/ and / ξ a/ elicited a characteristic MMNm-shape (Figure 3b). Visual inspection of the difference waves showed a MMNm-latency of about 170 ms both in case of /ba/ and / ξ a/. Accordingly, statistical analysis of the group data yielded a significantly increased difference wave at each hemisphere ($p < 0.02$). In contrast to the syllable /ba/ and the pseudo-syllable / ξ a/, the comparison of the deviant vowel /e/ with the standard /a/ resulted in difference wave peaking at about the same time as the N1m-component.

3.2. Experiment 2

Figure 4 shows the group averages of the elicited MMNm-amplitudes as determined 80 ms after the N1m-peak. In each subject, an overall increase of MMNm parallel to the lengthening of the F2-transient ($p < 0.05$) can be noted. Assuming this relationship to reflect a global linear progress reveals that stimuli comprising transients of a length of 30–40 ms elicited enlarged responses ($p < 0.05$). Only the fastest change in spectral energy distribution (transient length = 10 ms) gave rise to a significant lateralization effect ($p < 0.05$) in terms of a relatively left-emphasized MMNm.

4. DISCUSSION

Primarily, the present study aimed at the investigation of hemispheric lateralization of the processing of formant transients, varying either in direction or in duration, at the supratemporal plane. As concerns MMNm, only the most rapid change in spectral energy distribution (transient duration = 10 ms) elicited a significant lateralization effect, i.e., a larger amplitude over the dominant auditory cortex. The two cerebral hemispheres, thus, seem to differentially influence the encoding of auditory stimuli even at the level of pre-attentive mechanisms, as targeted by the mismatch design. However, this shortest F2-shift gives rise to the perception of a click noise rather than the consonant /b/. It is implausible, therefore, that the observed pre-attentive lateralization effect represents the neural correlate of the superior processing of consonantal sounds within the dominant perisylvian cortex, as reported by dichotic listening and clinical studies. The functional significance of the shortest F2-transient remains to be established. Conceivably, detection of these events mediates a first stage of the parsing of the acoustic stream conveyed to the ears [11].

The MMNm-amplitude increases across the prolongation of F2-transient length, i.e., the rise time of F2 up to the steady-state frequency of the template. Depending upon signal characteristics, two further components add to MMNm besides this cumulative spectral effect: the lateralized response due to transients of 10 ms and the bilateral enhancement evoked by shifts of 30 to 40 ms duration. Thus, MMNm may comprise various deviance reactions activated in parallel.

Provided that the interval between successive stimuli exceeds some short minimal time, any discrete auditory event reliably evokes long-latency responses comprising a negative EEG deflection peaking at about 100 ms (N1) after stimulus onset. This event-related potential (ERP) seems to reflect mere detection of abrupt stimulus changes with respect to, e.g., sound pressure or spectral features. MEG studies reported N1m-responses (magnetic equivalent of the EEG N1-component) to speech sounds, time-locked, e.g., to the onset of consonant-vowel sequences or to acoustic events within syllables such as the transition from a consonantal to the following vocalic segment. In accordance with these former findings, both the frequent and the deviants each elicited a N1m-component. This response failed any significant side differences both with respect to latency and amplitude. Thus, mere detection of the onset of isolated vowel or consonant stimuli does not seem to be mediated by a lateralized processing mechanism.

The difference wave in response to the vowel contrast considered, i.e., /a/ = frequent, /e/ = deviant event, showed shorter latency and higher amplitude than the CV-compound /ba/ and the pseudo-syllable / ξ a/. Diesch and Luce [12] obtained similar data and suggested that these effects mirror differences between syllable and vowel stimuli in terms of perceptual discriminability and auditory memory decay. As an alternative, the differential influence of vowel and consonant contrasts on the computed difference wave just might reflect adaptation processes. It is widely accepted that during peripheral auditory processing an array of overlapping bandpass filters acts upon acoustic stimuli and that signal components within the same domain (as a rule $\pm 20\%$ of center frequency) influence each other, e.g., in terms of masking

[13]. Whereas the acoustic energy of the two isolated vowels /a/ and /e/ are distributed across different channels, the transients of the /ba/- and / ξ a/-stimuli pass almost the same filters as the steady-state formants of vowel /a/, conceivably, giving rise to the reduced N1m-amplitudes.

ACKNOWLEDGEMENTS

This work has been supported by the *Deutsche Forschungsgemeinschaft* (Ac 55/5-1).

REFERENCES

- [1] E Zaidel. 1990. Language functions in the two hemispheres following complete cerebral commissurotomy and hemispherectomy. In F Boller and J Grafman (eds.), *Handbook of Neuropsychology*, volume 4, pages 115–150. Amsterdam: Elsevier.
- [2] M Studdert-Kennedy and D Shankweiler. 1970. Hemispheric specialization for speech perception. *Journal of the Acoustical Society of America*, 48:579–594.
- [3] A M Liberman. 1996. *Speech: A Special Code*. Cambridge, MA: MIT Press
- [4] P Tallal, S Miller, and R H Fitch. 1993. Neurobiological basis of speech: A case for the preeminence of temporal processing. *Annals of the New York Academy of Sciences*, 682:27–47.
- [5] P Tallal and F Newcombe. 1978. Impairment of auditory perception and language comprehension in dysphasia. *Brain and Language*, 5:13–24.
- [6] J Schwartz and P Tallal. 1980. Rate of acoustic change may underlie hemispheric specialization for speech perception. *Science*, 207:1380–1381.
- [7] M Steinschneider, J Arezzo, and H G Vaughan, Jr. 1982. Speech evoked activity in the auditory radiations and cortex of the awake monkey. *Brain Research*, 252:353–365.
- [8] J A Fiez, M E Raichle, F M Miezin, S E Petersen, P Tallal, and W F Katz. 1995. PET studies of auditory and phonological processing: Effects of stimulus characteristics and task demands. *Journal of Cognitive Neuroscience*, 7:357–375.
- [9] A C Maiste, A S Wiens, M J Hunt, M Scherg, and T W Picton. 1995. Event-related potentials and the categorical perception of speech sounds. *Ear and Hearing*, 16:68–90.
- [10] J E Cutting. 1974. Two left-hemisphere mechanisms in speech perception. *Perception and Psychophysics*, 16:601–612.
- [11] S M Boker and M Kubovy. 1998. The perception of segmentations in sequences: Local information provides the building block for global structure. In *Timing of Behavior: Neural, Psychological, and Computational Perspectives*, pages 109–123. Cambridge, MA: MIT Press.
- [12] E Diesch and T Luce. 1997. Magnetic mismatch fields elicited by vowels and consonants. *Experimental Brain Research*, 116:139–152.
- [13] B C J Moore. 1995. Frequency analysis and masking. In B C J Moore (ed.), *Hearing*, pages 161–205. San Diego: Academic Press.