MAGNETIC BRAIN ACTIVITY TRACING THE PERCEIVED SPEECH SIGNAL REGARDING ENVELOPE, SYLLABLE ONSETS, AND PITCH PERIODICITY

Ingo Hertrich, Susanne Dietrich, Jürgen Trouvain, Anja Moos & Hermann Ackermann

University of Tuebingen, Germany
ingo.hertrich@uni-tuebingen.de

ABSTRACT

Continuous speech evokes electrophysiological brain activity phase-locked to the speech envelope (ENV), resembling the N100 responses to single acoustic events. Using magnetoencophalography (MEG), the present study investigated further MEG components that directly reflect acoustic properties of continuous speech signals, i.e., a derivate of the envelope that can be taken as a physical marker of syllable onsets (SYL), and a pitch periodicity signal (PIT) obtained by bandpass filtering of the rectified speech signal. The participants (n = 10) listened to natural or formant-synthesized speech at a moderately fast or ultra-fast speaking rate. As expected, the MEG cross-correlation function with ENV showed a right-lateralized M100-like response with a source in the region of the central auditory system. Regarding the SYL derivate, in addition to auditory M50/M100-like responses a late, more anterior M50-like source component could be isolated. The cross-correlate of MEG data with the PIT derivate of the speech signal comprised multiple peaks of alternating polarity bound to a central-auditory source. The amplitude of the peaks in the cross-correlation functions depended upon speech rate and signal type (natural versus synthetic).

Keywords: speech perception, magnetoencephalography, cross-correlation

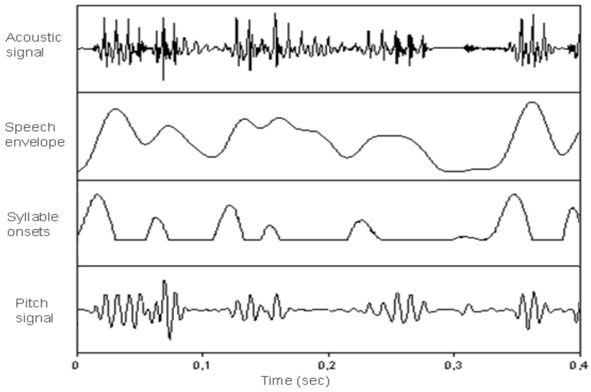
1. INTRODUCTION

Acoustic events evoke transient brain responses such as the electroencephalographic (EEG) N100 complex or its magnetoencephalographic (MEG) counterpart, the M100 field. Cross-correlation analysis of EEG signals with the speech envelope of continuous speech found a similar deflection. Since the speech envelope is closely related to the syllabic structure of speech, each syllable seems to evoke a brain activity similar to the one in response to single acoustic events. The amplitude of this electrophysiological correlate was stronger over the right as compared to the left hemisphere and was reduced in individuals with reading difficulties [1, 2]. Significant phase locking to the speech envelope has also been observed - though at a decreased amplitude - in response to time-compressed, unintelligible speech [3].

Besides the speech envelope, phase locking to other pivotal elements of the acoustic speech signal such as syllable onsets must be expected as well. Syllable onsets are, among others, relevant for the alignment of segmental (single speech sounds) and suprasegmental information units [8] and often characterized by a rapid increase of sonority [9].

Time locking between acoustic signals and electrophysiological responses has also been demonstrated within the range of higher modulation frequencies, e.g., in the 40-Hz region [5], and even up to frequencies about 100 Hz [7], i.e., well into the fundamental frequency (F0) domain of a male speaking voice.

Figure 1: Ultra-fast synthetized speech example (German): "Die Änderung wird, wie d..." The four panels (from top) show the oscillogram, the speech envelope, the positive part of its first derivative emphasizing syllable onsets, and the pitch periodicity signal.



Against this background, the present study investigates time-locking processes regarding three derivates of the acoustic speech signal, i.e., (i) the speech envelope (ENV), i.e. the intensity course of speech predominantly accounting for syllabic nuclei, (ii) the positive part of its first derivative, which could be taken as a crude indicator of the sonority rise following syllable onsets (SYL), and (iii) a pitch periodicity signal (PIT). In order to assess pitch-related phase locking, the acoustic speech signal was rectified and bandpass-filtered to obtain Fo as the instantaneous frequency of a sinusoidal trajectory representing amplitude modulation of higher frequency components. Examples of the three derivates of the speech signal are given in Figure 1.

The present study used two different speech rates, moderately fast (intelligible to normal subjects) and (largely unintelligible) ultra-fast. Furthermore, acoustic signal quality was varied by including natural as well as formant-synthesized speech. This latter aspect was motivated by a previous study showing that comprehension of ultra-fast speech by blind subjects is superior in case of formant synthesis as compared to the more natural-sounding diphone synthesis [10]. The applied cross-correlation method displays the amplitude and latency of MEG signal components that are directly correlated with the acoustic input. Furthermore, since the MEG cross-correlates represent time series similar to averaged MEG datasets, source localization procedures such as dipole analysis can be performed with these signals.

2. METHODS

2.1. Participants

Ten paid healthy subjects participated in the experiment. All of them were right-handed and showed unimpaired hearing thresholds.

2.2. Stimuli

Using a formant synthesizer (JAWS, 2008, male voice), 40 text passages were transformed into acoustic speech signals. In addition, a male human speaker read 40 further text passages. All utterances were recorded at a normal speaking rate of about 4-6 syl/sec and then compressed by means of the software package PRAAT [4]. One half of the material was re-synthesized with a moderately fast syllable rate (8 syl/sec) and the other half with ultra-fast rate (16 syl/sec). The total sample of 80 stimuli, thus, could be assigned to four combinations of the factors *rate* (moderately fast, ultra-fast) and *type* (spoken, formant-synthezied). A previous perceptual study, based upon the same test materials, had demonstrated that the moderately fast conditions were intelligible to normal subjects whereas ultra-fast speech was almost entirely incomprehensible (less than 10 percent correct words in a repetition task) apart from a blind subject included in that study [6].

2.3. Procedure

Subjects were seated within the MEG device (CTF, Canada, 272 channels, sampling rate = 585.938 Hz), the eyes being open. The entire experiment comprised 240 stimuli, i.e., three repetitions of each of the 80 stimuli. Onset-to-onset interstimulus interval amounted to 7 sec (4 sec speech followed by 3 sec pause).

2.4. MEG data processing

Three derivates of the speech signal were obtained by rectification and filtering in following way:

a) Speech envelope: The acoustic stimuli were bandpass filtered (80 - 2500 Hz), rectified and low-pass filtered at 25 Hz.

b) Syllable onsets: The first derivative of the speech envelope was computed, and all negative values of this signal were set to zero.

c) Pitch periodicity: The speech signal was bandpass-filtered (200 - 2300 Hz), concentrating on the spectral domain of the first two formants. Then the signal was rectified and subsequently bandpass-filtered (50 - 180 Hz) to the approximate range of a male speaking voice.

Prior to cross-correlation, the MEG sensor signals were bandpass-filtered (2-50 Hz for ENV and SYL, and 80-180 Hz for PIT). After onset alignment with the MEG data, the speech derivates were sample-by-sample shifted to the right on the time axis, and for each shift the signals were multiplied and added up in the following way:

$$\text{(1)} \qquad \mathrm{newsig}(t) = \sum_{j=1}^{n} \mathrm{sp}(j)\ \mathrm{MEGch}(t+j)$$

where sp is a derivate of the speech signal, MEGch is an MEG channel, t is the time shift between the two signals in sample points, and n is the length of the speech signal in samples. This operation was applied to all 272 MEG channels, resulting in a new MEG dataset that can be analyzed in a similar way as normal MEG data. In principle, this procedure can be considered as a kind of finite response filtering, with each speech derivate representing a set of filter coefficients. These coefficients were normalized to a sum of absolute values = 1, allowing for an estimation of the amplitude of signal-correlated MEG components.

Exploratory data analysis was performed at the level of group-averaged data to determine the latency and temporal extension of common signal components.

For statistical analysis, the strength of MEG field components was quantified using individual dipole models for the auditory M50/M100-like field, for the anterior M50 component, and for the pitch periodicity source.

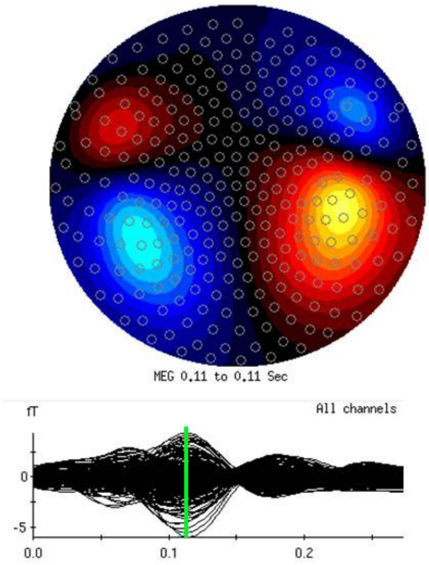
3. RESULTS

Cross-correlation of the MEG data with the speech events considered resulted in derived time courses, representing the temporal shift of the acoustic signal against the MEG data.

3.1. Group-averaged MEG data

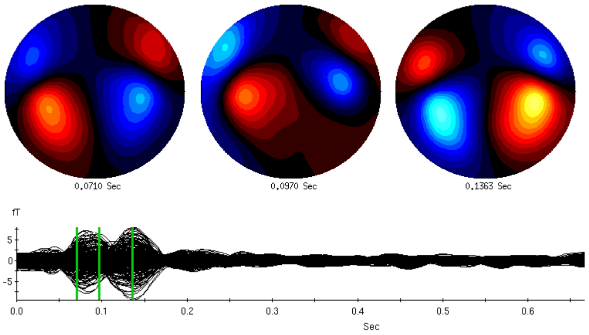
(i) The group average of the MEG cross-correlate with the speech envelope nicely resembled auditory evoked fields in response to single syllables. Figure 2 shows that the latency of the peak of global field power, displaying an M100-like field distribution, amounts to ca. 110 ms.

Figure 2: Cross-correlation with the speech envelope, upper panel: Surface map of the M100-like field at a latency of 110 ms, corresponding to the time of maximum field power in the lower panel (all channels).



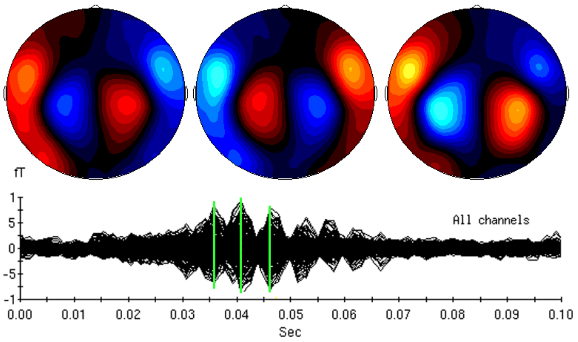
(ii) The syllable onsets derivate yielded a stronger and more complex response pattern. As shown in Figure 3, this response is characterized by an auditory M50-like field extending from ca. 70-100 ms. The field distribution continuously shifts within this time interval toward more anterior regions. Then the polarity changes toward the second large peak of global field power, displaying an auditory M100-like field distribution.

Figure 3: Upper panel: Surface maps of an early (left) and a late (mid) M50-like component followed by an M100-like field. The respective latencies are marked in the lower panel.



(iii) The pitch periodicity signal served as the third derivate of the speech materials. Its MEG cross-correlate is characterized by a series of amplitude peaks centered at a latency of ca. 41 ms. In spite of low absolute amplitude, this signal showed a good signal-to-noise ratio (Figure 4). The field distributions across these peaks display an alternation between an M50-like and an M100-like polarity.

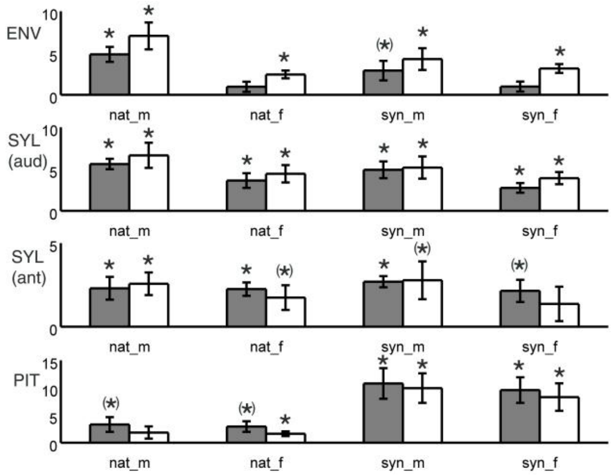
Figure 4: Upper panel: Surface maps of an early (left) and a late (mid) M50-like component followed by an M100-like field. The respective latencies are marked in the lower panel.



3.2. Statistical significance of MEG-Cross-correlates based on individual source models

Based on individual source models, time courses of dipole strength were rescaled in proportion to the magnitude of unspecific background activity estimated within a late time window (600-800 ms) of the respective cross-correlates. Figure 5 displays the strength of the peaks in the cross-correlation functions depending on conditions.

Figure 5: Group means and standard error of dipole strength in standardized units normalized to non-phase-locked background activity. ENV: M100-like field strength of the speech envelope (100-120 ms); SYL (aud): difference of M50- and M100-like deflections the syllable onset response; SYL (ant): strength of the anterior M50 component of the syllable onset response; PIT: Strength of the pitch periodicity-correlated MEG component. Significant values above zero: * p < 0.01, (*) p < 0.05 (two-tailed T-tests). Abbreviations: nat_m= natural, moderately fast; nat_f = natural, ultra-fast; syn_m = synthetic, moderately fast ;syn_f= synthetic, ultrafast.



4. DISCUSSION

In summary, the three speech derivates (envelope, syllable onsets, pitch) show differential patterns of speech-evoked MEG activity across the experimental conditions. As concerns the speech envelope, M100-like field deflections of a peak latency of about 110 ms could be observed, with a larger amplitude over the right as compared to the left hemisphere. A more complex pattern emerged in response to the sonority rises following syllable onsets, encompassing an earlier (upper temporal) and a later (frontal) M50-like structure preceding the M100 deflection. The MEG cross-correlate with pitch periodicity yielded a field pattern with an M50-like polarity at a latency of 41 ms, surrounded by additional peaks of alternating polarity, reflecting the fundamental frequency of a male voice.

5. REFERENCES

[1] Abrams, D.A., Nicol, T., Zecker, S., Kraus, N. 2008. Right-hemisphere auditory cortex is dominant for coding syllable patterns in speech. *J. Neuroscience* 28, 3958-3965.

[2] Abrams, D.A., Nicol, T., Zecker, S., Kraus, N. 2009. Abnormal Cortical Processing of the Syllable Rate of Speech in Poor Readers. *J. Neuroscience* 29, 7686-7693.

[3] Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H., Merzenich, M.M. 2001. Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proc. National Academy of Sciences* USA 98, 13367-13372.

[4] Boersma, P. Weenink, D. 2009. *Praat: doing phonetics by computer (Version 5.1.05)* [Computer program]. Retrieved May 1, 2009, from *http://www.praat.org/*.

[5] Gutschalk, A., Mase, R., Roth, R., Ille, N., Rupp, A., Scherg, M., et al. 1999. Deconvolution of 40 Hz steady-state fields reveals two overlapping source activities of the human auditory cortex. *Clinical Neurophysiology* 110, 856-868.

[6] Hertrich, I., Dietrich, S., Moos, A., Trouvain, J., Ackermann, H. 2009. Enhanced speech perception capabilities in a blind listener are associated with activation of fusiform gyrus and primary visual cortex. *Neurocase* 15, 163-170.

[7] Hertrich, I., Mathiak, K., Lutzenberger, W., Ackermann, H. 2004. Transient and phase-locked evoked magnetic fields in response to periodic acoustic signals. *Neuroreport* 15, 1687-1690.

[8] Ladd, D.R., Faulkner, D., Faulkner, H., Schepman, A. 1999. Constant ``segmental anchoring" of F[sub 0] movements under changes in speech rate. *J. Acoust. Soc. Am.* 106, 1543-1554.

[9] Leena, M., Yegnanarayana, B. 2008. Extraction and representation of prosodic features for language and speaker recognition. *Speech Communication* 50, 782-796.

[10] Trouvain, J. 2007. On the comprehension of extremely fast synthetic speech. *Saarland Working Papers in Linguistics* 1, 5-13.