Brain and Language 185 (2018) 19–29

Contents lists available at ScienceDirect

Brain and Language

journal homepage: www.elsevier.com/locate/b&l

Cortical phase locking to accelerated speech in blind and sighted listeners prior to and after training

Ingo Hertrich[a,*], Susanne Dietrich[b], Hermann Ackermann[a]

[a] *Department of Neurology, Hertie Institute for Clinical Brain Research, University of Tübingen, Germany*
[b] *Department of Psychology, Evolutionary Cognition (Cognitive Sciences), University of Tübingen, Germany*

ARTICLE INFO

Keywords:
Magnetoencephalography
Auditory processing
Speech perception
Blindness
Speech rate
Phase locking

ABSTRACT

Cross-correlation of magnetoencephalography (MEG) with time courses derived from the speech signal has shown differences in phase-locking between blind subjects able to comprehend accelerated speech and sighted controls. The present training study contributes to disentangle the effects of blindness and training. Both subject groups (baseline: n = 16 blind, 13 sighted; trained: 10 blind, 3 sighted) were able to enhance speech comprehension up to ca. 18 syllables per second. MEG responses phase-locked to syllable onsets were captured in five pre-defined source locations comprising left and right auditory cortex (A1), right visual cortex (V1), left inferior frontal gyrus (IFG) and left pre-supplementary motor area. Phase locking in A1 was consistently increased while V1 showed opposite training effects in blind and sighted subjects. Also the IFG showed some group differences indicating enhanced top-down strategies in sighted subjects while blind subjects may have a more fine-grained bottom-up resolution for accelerated speech.

1. Introduction

As a substitute for fast reading, blind individuals can develop the skill of understanding accelerated synthetic speech at syllable rates of up to about 20 syllables per second (syl/s) (Trouvain, 2007). Previous functional magnetic resonance imaging (fMRI) studies have shown that this ability is associated with significant hemodynamic activity of right visual cortex (Dietrich, Hertrich, & Ackermann, 2013b; Hertrich, Dietrich, Moos, Trouvain, & Ackermann, 2009). Based on dynamic causal modeling (DCM), a method of effective connectivity analysis, it was assumed that this activity is driven by subcortical auditory input via an audiovisual interface of the secondary visual pathway and further linked to the auditory cortex and the frontal speech processing network via the supplementary motor area (Dietrich, Hertrich, and Ackermann, 2015). Further support for these assumptions was provided by a diffusion tensor imaging study showing a significant positive correlations between the ability of ultra-fast speech comprehension and fractional anisotropy within right-hemisphere optic radiation and thalamus, experience-related structural alterations of brain connectivity (Dietrich, Hertrich, Kumar, and Ackermann, 2015). While secondary visual areas such as the face or object areas can be reorganized to engage in nonvisual processing in blind subjects via backward projections from supramodal representations (Röder, Ley, Shenoy, Kekunnaya, & Bottari, 2013), a study on anophthalmic subjects suggests that primary visual cortex may be activated by subcortical auditory input rather than backward projections (Watkins et al., 2012), which nicely corresponds to the above-mentioned findings on effective connectivity and auditory phase locking observed in a group of blind individuals.

Both in sighted and blind subjects, the demand of processing accelerated speech may elicit a listening strategy that engages top-down mechanisms of prediction and signal reconstruction. This is done by the speech generation modules in inferior frontal cortex, where hemodynamic activations have been observed in studies on the perception of fast speech (Poldrack et al., 2001; Vagharchakian, Dehaene-Lambertz, Pallier, & Dehaene, 2012). Furthermore, in case of high task demands the pre-supplementary motor area seems to have a particular control function with respect to top-down processing, first, with respect to its temporal coordination with the auditory input (Kotz, Schwartze, & Schmidt-Kassow, 2009) and, second, with respect to the inhibition of forward processing in case of errors, as has been shown, e.g., in case of stop signal tasks (Chao, Luo, Chang, & Li, 2009; Watanabe et al., 2015).

Phase-locking between the acoustic signal and brain activity can be assessed by means of magnetoencephalography (MEG) using a cross-correlation approach (Hertrich, Dietrich, Trouvain, Moos, & Ackermann, 2012). Thereby, the MEG data are cross-correlated with time courses derived from the acoustic signal such as the speech

---

[*] Corresponding author at: Center of Neurology, Hertie Institute for Clinical Brain Research, University of Tübingen, Hoppe-Seyler-Str. 3, D-72076 Tübingen, Germany.

E-mail address: ingo.hertrich@uni-tuebingen.de (I. Hertrich).

https://doi.org/10.1016/j.bandl.2018.07.002
Received 5 October 2017; Received in revised form 6 July 2018; Accepted 6 July 2018

envelope (bandwidth ca. 3–20 Hz) or a sinusoidal signal delineating pitch periodicity (ca. 65–130 Hz in case of a male voice). Such cross-correlation functions can be analyzed like auditory evoked fields, however, with the advantage that the stimulus material can be continuous speech and does not require multiple repetitions of identical or nearly-identical items for averaging. In this way, a previous study investigated MEG signal components phase-locked to the acoustic speech signal in blind subjects who were able to comprehend ultra-fast speech in comparison to a sighted control group without this ability (Hertrich, Dietrich, & Ackermann, 2013a). In addition to enhanced phase locking in auditory cortex, the blind subjects showed phase locking to syllable onsets in right visual cortex, suggesting that the blind subjects' ability to understand ultra-fast speech is associated with changes in early signal-related processing mechanisms both within and outside the central-auditory system.

The present study can be considered as an extension of Hertrich et al. (2013a) to disentangle the effects of blindness and learning. To these ends, blind and sighted subjects were behaviorally tested (sentence repetition task) and MEG-recorded (during listening to speech) prior to and after a training period in which they had to acquire the skill of understanding accelerated synthetic speech. The analysis of the MEG data relied on a hypothesis-driven approach considering pre-defined virtual source locations for testing the strength of phase locking. These locations were taken from the fMRI study of Dietrich et al. (2013b) showing fMRI activation clusters with significant covariance with the ability to understand ultra-fast speech. Based on a hypothesis of right-hemispheric tracking of syllabic events (Hertrich, Dietrich, & Ackermann, 2013b), three sources were considered, comprising right visual cortex (V1), left pre-supplementary motor area (pre-SMA), and left inferior frontal cortex (IFG). Additionally, based on anatomical data, left and right Heschl's gyri were also included in the source model in order to capture the bulk of auditory evoked activity. As in Hertrich et al. (2013a), reversed acoustic signals were used as unintelligible control stimuli.

It was hypothesized that the training procedure generally enhances the phase locking between MEG data and the acoustic signal in auditory and frontal cortex whereas phase locking in visual cortex was particularly expected in trained blind subjects during the perception of accelerated speech. The pre-SMA, with subcortical input from the basal ganglia and the cerebellum and (mostly inhibitory) output via the frontal Aslant tract toward the inferior frontal language areas, can be considered as an area of cognitive control upon speech processing. So far, however, it is not clear to what extent the activity of pre-SMA is phase-locked to the speech signal.

2. Materials and methods

2.1. Subjects and training procedure

Overall, the study included 13 sighted (6 female; mean age = 36.5 years, SD = 11.9) and 16 blind or severely vision-impaired subjects (7 female; mean age = 40.2 years, SD = 13.1; in the following labeled as “blind”). The study was approved by the Ethics Committee of the University of Tübingen, written consent was obtained from all subjects in line with the guidelines of the Ethics Committee, and the subjects were paid for their participation including additional compensation for travelling costs. Blind subjects were recruited from blind communities and support groups and by advertisement in a magazine. Blindness was acquired during childhood (11 cases) or adult stages (5), none of the subjects were blind from birth. Medical diagnosis was retinitis pigmentosa (8), macular degeneration (2), extreme myopia (2), optic nerve tumor (1), glaucoma (1), iris coloboma (1) or cataract (1). All subjects participated in an initial behavioral test and a subsequent MEG session. Additionally, three of the sighted subjects and all blind subjects were required to participate in a training procedure. Using the text-to-speech screen reader “JAWS” (Freedom Scientific; formant synthesizer “eloquence”, implemented on a laptop) for regular electronic newspaper reception during a period of some months, they had to continuously increase the syllable rate of the synthetic speech output until they felt able to approximately comprehend speech at 13 syl/s which was defined as the first training stage, followed by behavioral testing and a second MEG session. Then they had to continue the training until they could nearly cope with a syllable rate of 18 syl/s, i.e., the stage for final testing and MEG recording. Due to some drop-outs, only 10 blind and the three sighted subjects participated in the third session. Furthermore, only seven of the blind subjects reached the originally intended target of understanding more than 50% of the speech material at the third session. In total, 58 MEG sessions could be evaluated, comprising 29 baseline recordings, 16 recordings at the second stage, and 13 recordings at the third stage.

2.2. Behavioral testing

Prior to each MEG session, the subjects' ability to understand accelerated speech was tested by means of a sentence repetition task comprising 33 sentences of 18 syllables each. All sentences were text-to-speech synthesized at 9 different syllable rates ranging from 6 to 22 syl/s using the same synthesizer (JAWS) as for the training procedure. In order to avoid a large number of trials with ceiling or bottom effects, a staircase-like procedure was applied to determine the percentage of intelligibility as a function of syllable rate without an excessive amount of trials. The test started with a slow syllable rate of 6 syl/s. Speech rate was then increased by 2 syl/s in subsequent stimuli up to the upper limit of 22 syl/s. Then the syllable rate was reduced by 2 syl/s down to a syllable rate at which subjects could understand more than ca. 50% of the linguistic material, then it was increased again until performance was less than 50% and so on. For evaluation, the percentage of correct syllables was plotted against syllable rate, and a psychometric function was fitted to the obtained data. From this function, the percentage of correct syllables at a rate of 18 syl/s was determined as a parameter of performance.

2.3. Stimuli of the MEG experiment

The speech stimuli for the MEG experiment were similar to the ones used in Dietrich, Hertrich, & Ackermann (2013a): Three sets (to be used at different training stages, balanced across subjects) of 90 different text passages (comprising one or two sentences each) were collected from newspapers or similar sources and converted to speech (using JAWS, see above) at different speaking rates (30 stimuli each at 8, 13, or 18 syl/s). Duration was ca. 4 s for all stimuli. Additionally, all stimuli were converted into time-reversed speech signals, serving as spectrally matched unintelligible control items. Altogether, each stimulus set contained 30 (items) × 3 (rates) × 2 (directions: forward/backward) = 180 stimuli that were presented in pseudo-randomized order within one session, subdivided into 5 runs with equal distributions of syllable rate and direction.

2.4. Procedure of MEG recording

Subjects were seated within the MEG device (CTF, Canada, 272 channels, sampling rate = 586 Hz) with their eyes open, directed at a fixation cross in front of them (for sighted subjects). The speech stimuli were played via MEG-compatible air-tube headphones, and the subjects just had to listen attentively without any additional behavioral task. The entire session comprised five runs of 36 stimuli each, presented at an onset-to-onset inter-stimulus interval of 7 sec (4 sec stimulus followed by 3 sec pause). The MEG was recorded continuously, together with a trigger channel for the alignment of the stimulus onsets with the MEG data.