Perception of Ultra-Fast Speech by a Blind Listener – Does He Use His Visual System?

Anja Moos*, Ingo Hertrich°, Susanne Dietrich°, Jürgen Trouvain*, Hermann Ackermann°

Saarland University, °University of Tübingen

E-mail: {anmo|trouvain}@coli.uni-saarland.de, ingo.hertrich@uni-tuebingen.de

Abstract

We investigated if a blind person uses his visual system to understand moderately fast speech (8 syll/sec) and ultra-fast speech (16 syll/sec) based on texts spoken by a male person and produced by a speech synthesis, respectively. Whereas the blind subject had no problems understanding ultra-fast speech, six sighted control subjects were not able to understand it. Functional magnetic resonance imaging (fMRI) of the brain activity proved that moderately fast speech activated posterior and anterior 'language zones' in all subjects. Regarding ultra-fast tokens, the controls showed exclusive activation of supratemporal regions whereas the blind participant exhibited enhanced left inferior frontal and temporoparietal responses as well as significant hemodynamic activation of left fusiform gyrus (FG) and right primary visual cortex. Since left FG is known to be involved in phonological processing, this structure, presumably, provides the functional link between the central-auditory and -visual systems.

1 Introduction

Speech perception has often been considered as audiovisual in nature, and the visual system seems to be connected to the auditory system in various ways. Thus, it seems plausible that blind people are able to use parts of their visual system to enhance speech processing. Previous studies have shown blindness-related improvements in lexical-semantic and verbal memory tasks [1, 13]. Furthermore, the functional relevance of visual cortex activation for language processing in blind subjects has been explicitly demonstrated by means of repetitive transcranial magnetic stimulation [2].

Some blind listeners are able to perceive ultra-fast synthetic speech exceeding 20 syll/sec, which is far beyond the maximum performance (ca. 10 syll/sec) of sighted listeners [9] as well as the maximal articulation rate in speech production (ca. 8-12 syll/sec). The present single-case study is a first approach to assess the brain activity associated with this exceptional ability by performing a functional magnetic resonance imaging (fMRI) experiment.

2 Methods

2.1 Subjects

A blind person (male, 26 years, blind for 14 years, very experienced with speech synthesis) and six normal-sighted control participants (mean age = 28 years, SD = 4 years; 2 females, 4 males; no experience with speech synthesis) took part in this study. All subjects were German native speakers.

2.2 Material

The stimulus material (extracted from newspaper reports) comprised 80 text portions of 28-32 syllables for the fast and 57-64 syllables for the ultra-fast stimuli, so that each stimulus had a length of 3.5 – 4 sec. 40 items were produced by a male speaker and 40 were generated by formant synthesis (male voice) integrated in the screen reader software JAWS [8]. All sentences were recorded in a normal speaking rate of about 4-6 syll/sec and then compressed in Praat [10] using the PSOLA method [6]: The fundamental frequency remains unchanged while the durations are changed linearly by averaging over-lapping parts of adjacent F0 periods in the time domain. After compression, the material comprised 20 moderately fast and 20 ultra-fast stimuli for each

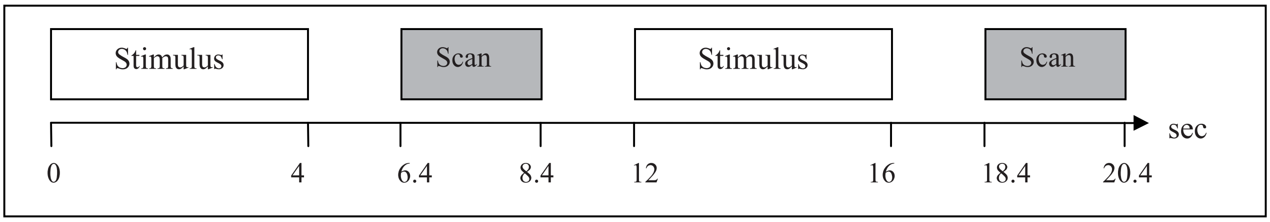

Figure 1: *Timing (sec) of the fMRI recordings: Each stimulus is followed by a 2 sec whole-head scan capturing the peak phase of the bold response. Interstimulus interval = 12 sec.*

speech mode: synthesized and naturally spoken utterances. This results in four stimulus conditions: 'syn_16', 'syn_8', 'nat_16' and 'nat_8'. 'Fast' and 'ultra-fast' corresponds to an articulation rate of 8 syll/sec and 16 syll/sec, respectively.

2.3 fMRI data acquisition

Each fMRI session encompassed five runs separated by short breaks. Each run comprised 40 epochs of 12 sec (pseudo-randomized presentation of 32 speech stimuli and 8 silent baseline intervals as empty control condition). Thus, all the 80 different speech stimuli were applied twice, the two repetitions pertaining to different runs each. Functional imaging was conducted on a 3-T Siemens scanner (TRIO; Siemens) using a sparse sampling design (see Fig. 1).

2.4 Verification of comprehension

A short repetition task, comprising 10-word utterances extracted from the speech material used in the scanner, was performed after the fMRI session in order to verify that the blind listener, in contrast to the controls, was indeed able to understand the ultra-fast synthetic passages.

3 Results

The repetition task showed that the behavorial performance of the blind subject, in terms of the percentage of correctly repeated words, was within the normal range (see Fig. 2). By contrast, his understanding of ultra-fast speech by far exceeded that of each single control subject. Furthermore, and in accordance with our previous study [9], recognition rate for ultra-fast speech tended to be higher in the case of synthetic as compared to accelerated natural stimuli.

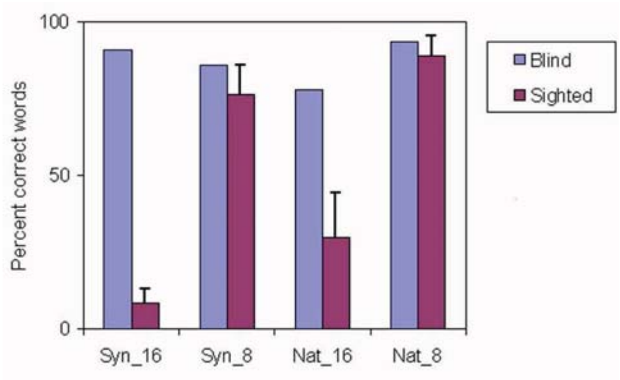

Figure 2: Percentage of correctly repeated words based on a repetition task comprising 10 utterances of ca. 10 words for each of the four conditions. Error bars indicate SD of the control group.

The fMRI data of each subject were analyzed separately, contrasting the empty control condition with all four stimulus conditions, as exemplified in Figure 3 where those areas are highlighted which are activated in the speech stimulus conditions (in contrast to the empty control conditions).

(i) All four experimental conditions gave rise to bilateral hemodynamic activation clusters within the temporal lobes, more or less encroaching upon temporoparietal regions in the left hemisphere. In the blind participant, the left-hemisphere cluster tended to be larger in response to ultra-fast as compared to moderately fast speech whereas all six controls showed the reverse pattern, i.e., decreased cluster size during the ultra-fast condition.

(ii) Under all conditions, the blind participant exhibited hemodynamic responses of left inferior frontal cortex (Brodmann area [BA] 44 and 45) as well as left precentral gyrus (PcG) and right cerebellum. By contrast, the controls showed this activation pattern exclusively during application of moderately fast speech. As concerns ultra-fast speech, hemodynamic activation was restricted to the temporal lobe in normal-sighted subjects.

(iii) In contrast to any of the control participants, most noteworthy, the blind listener showed significant hemodynamic activation of right occipital cortex (posterior parts of BA 17/18) and left fusiform gyrus (FG, lower part of BA 37) while listening to ultra-fast speech.

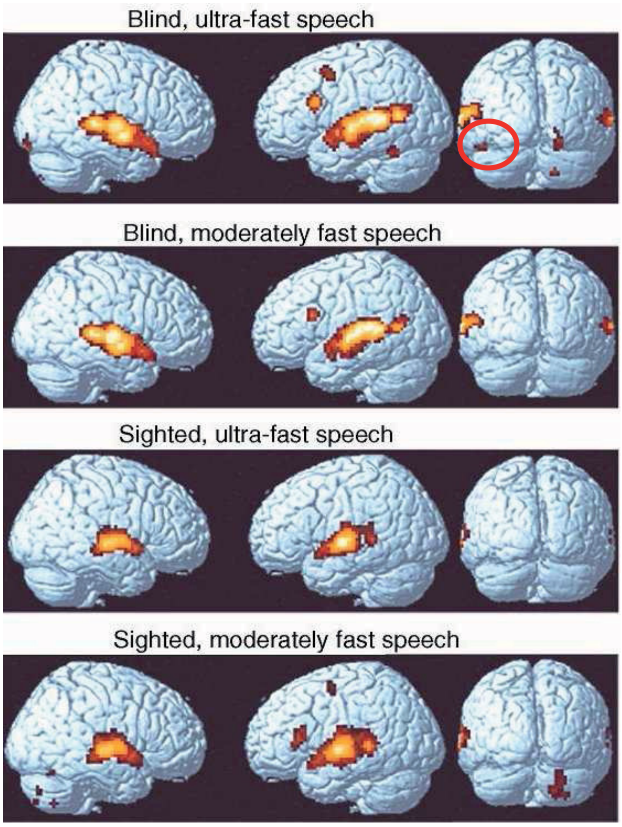

Figure 3: Activation versus baseline (empty control condition) in response to ultra-fast and moderately fast synthetic speech, shown for the blind subject and a typical sighted control subject: right hemisphere (left), left hemisphere (middle), dorsal view (right). Left FG in red circle for the blind at ultra-fast speech.

4 Discussion

As the most important finding of this study, the cerebral network bound to auditory speech processing was found to encroach upon the primary visual cortex in a blind participant being capable of understanding ultra-fast natural and synthetic speech. These effects did not emerge in any of the normal-sighted subjects. The observed interaction of blindness with comprehension of speeded speech at the level of left-hemisphere temporoparietal and frontal areas, presumably, reflects a 'dose effect': The blind subject successfully processed approximately twice as much linguistic information per time interval in case of ultra-fast as compared to moderately fast speech whereas, by contrast, the control subjects perceived fewer units during listening to ultra-fast speech.

There is first evidence that blind subjects are able to engage their visual system in order to enhance the sequencing of non-speech auditory stimuli. For example, a recent electrophysiological study, addressing attention-dependent processing of temporal versus spatial auditory stimulus properties, demonstrated improved temporal resolution capabilities in blind individuals [14]. In addition an fMRI experiment based upon temporal order judgements of backward-masked tone stimuli, found blind individuals to outperform a control group, particularly when the masker occurred at a brief interval (40 ms) after the tone sequence [15]. Most noteworthy, this condition is approximately comparable to the ultra-fast speech condition of the present study, given that each syllable acts as a potential masker of the preceding one. The enhanced performance of the blind subjects in [15] was accompanied by two activation loci within the primary visual cortex. The more posterior one of these responses nicely corresponds to the occipital activation spot of the present results. If these findings from experiments based on non-speech stimuli also hold for the domain of spoken language processing, accelerated extraction of linguistic information can be expected in blind subjects.

Besides primary visual cortex, listening to ultra-fast speech elicited a hemodynamic response of left FG in the blind participant of this study. This region has been found to be engaged in phonological operations [3, 7]. Furthermore, impaired phonological processing in children with reading difficulties is associated with reduced connectivity between FG and inferior parietal as well as frontal language areas [5]. Other studies were able to demonstrate an auditory timing deficit related to their phonological disorder in at least a subgroup of these children [12]. As compared to these dyslexic subjects, the blind listener showed the opposite effect: He was able to increase the speed of auditory speech processing. Given the fMRI data of the present study, left FG appears to link the processing capabilities of the occipital cortex – in terms of enhanced resolution of auditory signals in blind subjects (see above) – to the perisylvian language area. The observed contralaterality pattern of hemodynamic activation, i.e., left FG together with the right posterior occipital cortex, might be related to the blind subject's use of his left index finger during Braille reading (see [4] for a discussion on contralaterality effects in braille readers). In line with studies addressing other aspects of audio-visual interactions such as e.g. visual motion [11], the present results confirm the assumption that occipital recruitment in our blind subject is mediated by the same functionally specific cross-modal connections as in sighted subjects.

Further research with more than just a single subject is needed to show whether the findings presented here can be generalized. In addition it is interesting to find out whether and to which degree factors such as congenital vs. acquired blindness, training with time-scaled synthetic speech, and mode of speech (formant synthesis vs. samples based on natural speech) play a role for the usage of the visual system for speech decoding.

References

[1] A. Amedi, N. Raz, P. Pianka, R. Malach & E. Zohary. Early 'visual' cortex activation correlates with superior verbal memory performance in the blind. *Nature Neuroscience*, 6: 758-766, 2003.

[2] A. Amedi, A. Floel, S. Knecht, E. Zohary & L.G. Cohen. Transcranial magnetic stimulation of the occipital pole interferes with verbal processing in blind subjects. *Nature Neuroscience*, 7: 1266-1270, 2004.

[3] J.R. Binder, D.A. Medler, C.F. Westbury, E. Liebenthal & L. Buchanan. Tuning of the human left fusiform gyrus to sublexical orthographic structure. *Neuroimage, 33*: 739-74, 2006.

[4] H. Burton, A.Z. Snyder, T.E. Conturo, E. Akbudak, J.M. Ollinger & M.E. Raichle. Adaptive Changes in Early and Late Blind: A fMRI Study of Braille Reading. *Journal of Neurophysiology, 87*: 589-607, 2002.

[5] F. Cao, T. Bitan, & J.R. Booth. Effective brain connectivity in children with reading difficulties during phonological processing. *Brain and Language, in press*, corrected proof.

[6] F. Charpentier & E. Moulines. Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Proc. 2nd Eurospeech*: 13-19, 1989.

[7] N.A.E. Dietz, K.M. Jones, L. Gareau, T.A. Zeffiro & G.F. Eden. Phonological decoding involves left posterior fusiform gyrus. *Human Brain Mapping, 26:* 81-93, 2005.

[8] JAWS (Job Access With Speech) Screenreader software, http://www.freedomsci.de, visited 21-Jan-06.

[9] A. Moos & J. Trouvain. Comprehension of ultra-fast speech – blind vs. "normally hearing" persons. Proc. 16th *International Congress of Phonetic Sciences*: 677-680, 2007.

[10] PRAAT version 4.5, http://www.fon.hum.uva.nl/praat, visited 10-Jan-07.

[11] C. Poirier, O. Collignon, C. Scheiber, L. Renier, A. Vanlierde, D. Tranduy, C. Veraart, & A.G. De Volder. Auditory motion perception activates visual motion areas in early blind subjects. *Neuroimage, 31*: 279-285, 2006.

[12] F. Ramus. Developmental dyslexia: specific phonological deficit or general sensorimotor dysfunction? *Current Opinion in Neurobiology, 13:* 212-218, 2003.

[13] B. Röder, F. Rösler & H.J. Neville. Auditory memory in congenitally blind adults: a behavioral-electrophysiological investigation. *Cognitive Brain Research*, 11: 289-303, 2001.

[14] B. Röder, U.M. Kramer & K. Lange. Congenitally blind humans use different stimulus selection strategies in hearing: An ERP study of spatial and temporal attention. *Restorative Neurology and Neuroscience, 25*: 311-322, 2007.

[15] A.A. Stevens, M. Snodgrass, D. Schwartz & K. Weaver. Preparatory activity in occipital cortex in early blind humans predicts auditory perceptual performance. *Journal of Neuroscience*, 27: 10734-10741, 2007.