NeuroImage 49 (2010) 2697–2706

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Comprehension of degraded speech sounds with m-sequence modulation: An fMRI study ☆

Hiroshige Takeichi [a,b,c,*], Sachiko Koyama [c], Atsushi Terao [d], Fumiya Takeuchi [c,e], Yuko Toyosawa [c], Harumitsu Murohashi [c,f]

[a] Laboratory for Mathematical Neuroscience, Brain Science Institute, RIKEN, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan
[b] Intelligent Modeling Laboratory, the University of Tokyo, Tokyo, Japan
[c] RISTEX, JST, Tokyo, Japan
[d] School of Social Informatics, Aoyama Gakuin University, Sagamihara, Japan
[e] Faculty of Health Sciences, Hokkaido University, Sapporo, Japan
[f] Faculty of Education, Hokkaido University, Sapporo, Japan

ARTICLE INFO

Article history:
Received 9 January 2009
Revised 16 October 2009
Accepted 21 October 2009
Available online 28 October 2009

Keywords:
Inferior frontal gyrus
Premotor area
Middle temporal gyrus
Angular gyrus
Degraded speech
Japanese
SPM

ABSTRACT

In a recent electroencephalography (EEG) study (Takeichi et al., 2007a), we developed a new technique for assessing speech comprehension using speech degraded by m-sequence modulation and found a correlation peak with a 400-ms delay. This peak depended on the comprehensibility of the modulated speech sounds. Here we report the results of a functional magnetic resonance imaging (fMRI) experiment comparable to our previous EEG experiment. We examined brain areas related to verbal comprehension of the modulated speech sound to examine which neural system processes this modulated speech. A non-integer, alternating-block factorial design was used with 23 Japanese-speaking participants, with time reversal and m-sequence modulation as factors. A main effect of time reversal was found in the left temporal cortex along the superior temporal sulcus (BA21 and BA39), left precentral gyrus (BA6) and right inferior temporal gyrus (BA21). A main effect of modulation was found in the left postcentral gyrus (BA43) and the right medial frontal gyri (BA6) as an increase by modulation and in the left temporal cortex (BA21, 39), parahippocampal gyrus (BA34), posterior cingulate (BA23), caudate and thalamus and right superior temporal gyrus (BA38) as a decrease by modulation. An interaction effect associated specifically with non-modulated speech was found in the left frontal gyrus (BA47), left occipital cortex in the cuneus (BA18), left precuneus (BA7, 31), right precuneus (BA31) and right thalamus (forward > reverse). The other interaction effect associated specifically with modulation of speech sound was found in the inferior frontal gyrus in the opercular area (BA44) (forward > reverse). Estimated scalp projection of the component correlation function (Cao et al., 2002) for the corresponding EEG data (Takeichi et al., 2007a, showed leftward dominance. Hence, activities in the superior temporal sulcus (BA21 and BA39), which are commonly observed for speech processing, as well as left precentral gyrus (BA6) and left inferior frontal gyrus in the opercular area (BA44) is suggested to contribute to the comprehension-related EEG signal.

Language processing includes phonemic, lexical, syntactic and semantic processing. Verbal comprehension is achieved through interactive integrations among these processes, which Hagoort (2005) has called "unification" in relation to psycholinguistic models (e.g. Jackendoff, 2002; Chomsky, 1995). Multiple feed-forward and feed-back flows of information would be expected to occur between the processing levels. For example, a cohort model of verbal comprehension postulates that while bottom–up phoneme processing dominates at initial stages of sentence processing, top–down lexico-semantic processing dominates at later stages (Marslen-Wilson and Tyler, 1980; Marslen-Wilson, 1987).

Studies using event-related or evoked brain potentials (ERPs) have significantly contributed to our understanding of language processing (e.g. Kutas and Federmeier, 2000; Pulvermuller and Shtyrov, 2006). Conventional ERP measurements have two major limitations, however. First, it is difficult to extract ERP components when the process in question, for example verbal comprehension, does not have an externally well-defined onset time. In other words, triggers for ERPs cannot be clearly placed in natural verbal comprehension. Second, due to the requirement of an averaging procedure, a stimulus needs to be presented repeatedly in a long experimental session (often more than an hour) to separate ERPs from spontaneous background activity. To

☆ Supported by RISTEX-JST (Brain and Education to SK), KAKENHI (20530674 to HT and 19203038 to HM), and Nishina Center for Accelerator Based Research, RIKEN.

* Corresponding author. Laboratory for Mathematical Neuroscience, Brain Science Institute, RIKEN, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan. Fax: +81 48 467 9693.
E-mail address: takeichi@riken.jp (H. Takeichi).

1053-8119/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2009.10.063

overcome these problems, we developed a novel technique using m-sequence (maximum length shift register code) modulation (Takeichi et al., 2007a). We originally developed this technique to objectively assess the unifying process of verbal comprehension in a short time, independent of syntactic or phonemic performance and vocabulary.

Using this technique, we first replaced parts of the continuous speech sound stimulus with short temporal gaps of variable durations (see Fig. 1) according to a binary m-sequence. The m-sequence is a pseudorandom binary number (0 or 1) sequence generated on the basis of a linear recurrence equation $X_n = X_{n-p}$ XOR X_{n-q}, where p and q determine its cycle length and characteristics and XOR denotes an exclusive OR. The m-sequence has a broad bandwidth and is therefore suited to handling a variety of signal encoding operations under noisy environments. The validity of the m-sequence modulation has been shown using a variety of physiological measurements, such as retinal responses (Marmarelis and Naka, 1972; Sutter, 2001), otoacoustic responses (Thornton, 1993a,b), receptive field mapping (Reid et al., 1997) and responses from human visual cortex (Hansen et al., 2004; Fukunaga et al., 2004; Baker et al., 2006; Momose, 2007). In the context of speech comprehension, we assumed that signals acquired during and after each of the silent intervals (gaps) correspond to neural computations used to "bridge" the gaps in the stimuli for comprehension of the narrative, depending on the availability of contextual information. In previous studies, we recorded both electroencephalography (EEG) and magnetoencephalography (MEG) data while each subject was listening to modulated speech. Time-locked changes elicited by the modulated speech in the EEG signals were detected by computing a cross-correlation function (Matani et al., 2005) between the modulator m-sequence and the EEG signals. Independent component analysis (ICA) was then applied to the correlation functions to examine components specific to verbal comprehension. As a result, we identified a cross-correlation peak (Takeichi et al., 2007a,b; Takeichi et al., 2008a; for MEG results, see also Takeuchi et al., 2008), which was observed with a 400-ms delay between the m-sequence and the EEG record, exclusively for the comprehensible speech sound. Incomprehensible time reversal or foreign speech sounds did not induce this peak. Thus, this protocol has potentially provided an objective measure of verbal comprehension.

Neuroimaging studies have been updating the notion of a *language area* (see Hickok and Poeppel, 2007). Scott et al. (2000) and Obleser et al. (2007) used noise vocoding stimuli (Shannon et al., 1995; see also Riquimaroux 2006) in which speech sounds were replaced with noise, but the intensity envelope was kept intact. They found that anterior and ventral parts of the superior temporal sulcus were activated only by intelligible speech. The purpose of the present fMRI study was to determine the brain areas related to verbal comprehension of speech

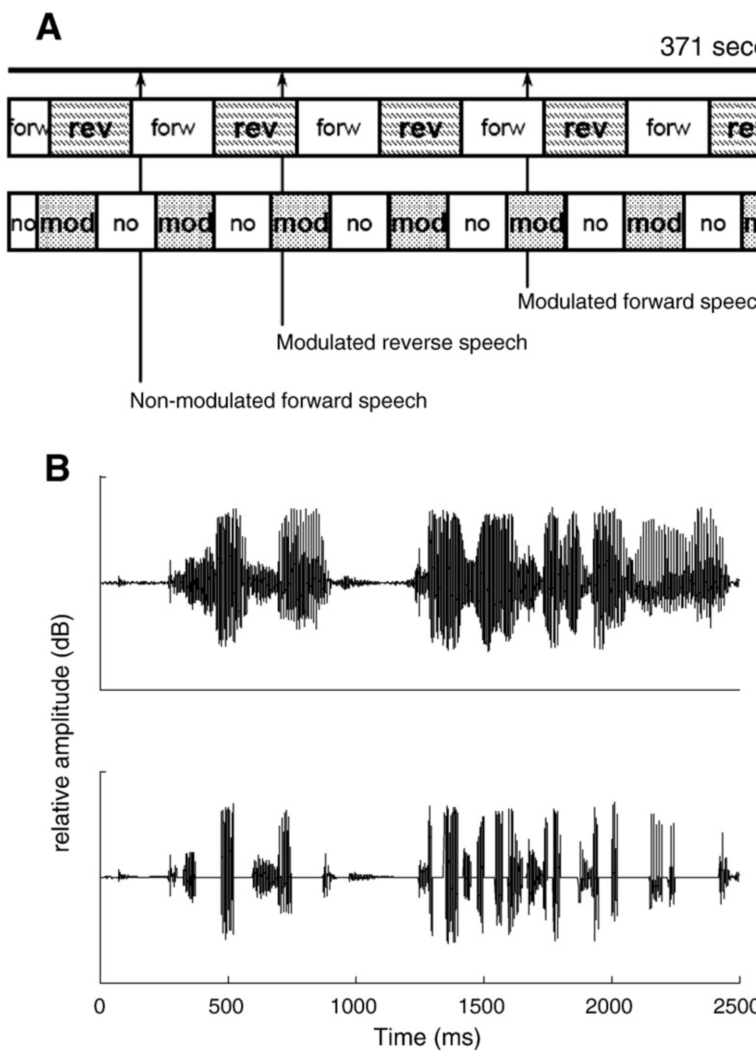

Fig. 1. (A) A schematic representation of the stimulus. Each of the 371-s stimuli was partially reversed in time at a frequency of five cycles per scan and partially and concurrently modulated by m-sequence at a frequency of seven cycles per scan. (B) Waveforms of a part of the Japanese forward sound and of the modulated sound. The speech here shows a male voice saying "sei-fu-wa koi-zu-mi-sou-ri-dai-jin-ga" literally meaning "(in reference to) Prime Minister Koizumi, the government (announced that)." Despite the modulation gaps that are sparsely introduced, the modulated speech is still comprehensible by itself.