Neuropsychologia 62 (2014) 233–244

Contents lists available at ScienceDirect

Neuropsychologia

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

Voice reinstatement modulates neural indices of continuous word recognition

Sandra Campeanu [a,b,*], Fergus I.M. Craik [a,b], Kristina C. Backer [a,b], Claude Alain [a,b]

[a] Rotman Research Institute, Baycrest Centre for Geriatric Care, Toronto, Ontario, Canada M6A 2E1
[b] Department of Psychology, University of Toronto, Ontario, Canada M5S 3G3

ARTICLE INFO

Article history:
Received 19 January 2014
Received in revised form
4 July 2014
Accepted 19 July 2014
Available online 29 July 2014

Keywords:
Auditory memory
Voice
ERP
Recognition memory
Source memory

ABSTRACT

The present study was designed to examine listeners' ability to use voice information incidentally during spoken word recognition. We recorded event-related brain potentials (ERPs) during a continuous recognition paradigm in which participants indicated on each trial whether the spoken word was "new" or "old." Old items were presented at 2, 8 or 16 words following the first presentation. Context congruency was manipulated by having the same word repeated by either the same speaker or a different speaker. The different speaker could share the gender, accent or neither feature with the word presented the first time. Participants' accuracy was greatest when the old word was spoken by the same speaker than by a different speaker. In addition, accuracy decreased with increasing lag. The correct identification of old words was accompanied by an enhanced late positivity over parietal sites, with no difference found between voice congruency conditions. In contrast, an earlier voice reinstatement effect was observed over frontal sites, an index of priming that preceded recollection in this task. Our results provide further evidence that acoustic and semantic information are integrated into a unified trace and that acoustic information facilitates spoken word recollection.

1. Introduction

Context congruency facilitates recognition memory (Tulving & Thomson, 1973; Morris Bransford & Franks, 1977; Craik & Kirsner, 1974; Palmeri, Goldinger & Pisoni, 1993; Bradlow, Nygaard & Pisoni, 1999; Goldinger, 1996; Campeanu, Craik, & Alain, 2013). In memory research, context refers to any number of details related to the initial study episode: location, font, voice and testing conditions are a few examples. In the auditory domain, perhaps the most important contextual cue for spoken word recognition is voice. A review by Pisoni (1993) proposed that speaker voice is encoded in a detailed perceptual representation with the episode itself. One question that emerges, then, is whether voice congruency between study and test acts to improve recognition memory in an all-or-none or in a graded fashion. For instance, voice information may be a gestalt-like "tag" with the word, such that a benefit of voice congruency occurs only for the same speaker. However, another possibility is that voice facilitates spoken word memory via a sense of familiarity, thereby making a partial congruency effect possible.

In a recent study, Campeanu et al. (2013) investigated the neural correlates of voice reinstatement on spoken word memory. Participants were presented with lists of words and were instructed to remember the word as well as the voice of the speaker. During the test phases, participants indicated whether the word was old or new and whether the old words were spoken by the same or a different speaker between study and test. There were four different speakers, resulting in a combination of gender and accent (Chinese or Canadian) congruency. Participants were more accurate in recognizing old words spoken by the same speaker during study and test blocks (i.e., a same-speaker voice reinstatement benefit). There was also an indication that accent congruency conferred a partial word recognition benefit.

In the present study, we used a continuous recognition paradigm to assess the impact of voice congruency on spoken word recognition. We sought to extend the characterization of a voice congruency effect by using a different paradigm than that used by Campeanu et al. (2013), but one that also allows a direct comparison with previous behavioral research (Craik & Kirsner, 1974; Palmeri et al., 1993). While our previous investigation used a block design (Campeanu et al., 2013), and emphasized attention to words and to voice at study (since both were directly tested), the current experiment asked participants to make word judgments alone, with no instruction regarding attention allocation.

With respect to voice congruency, the second presentation of each word may be spoken by the same voice, a voice of different gender but the same accent, a voice of the same gender but different accent, or a voice of different gender and different accent.

* Corresponding author. Tel.: +1 416 785 2500x3109; fax: +1 416 785 2862.
E-mail address: scampeanu@research.baycrest.org (S. Campeanu).

http://dx.doi.org/10.1016/j.neuropsychologia.2014.07.022
0028-3932/

We anticipated overall superior performance for old words spoken in the same voice. Moreover, though voice effects might be expected to decrease over time or with the number of intervening items, a complete absence of a same-voice advantage would be expected only at delays exceeding one day (Goldinger, 1996), a far more extended time course than we investigate in the present experiment. In addition, we predicted overall memory performance to be inversely correlated with the number of items and time between the first and second presentation (i.e., lag).

Prior studies (e.g., Wilding, Doyle, & Rugg 1995; Wilding & Rugg, 1996; Tendolkar et al. 2004), including some using a continuous recognition paradigm (i.e., Nielsen-Bohlman & Knight, 1994), have revealed a late positivity at parietal sites that is thought to index recollective processes. This modulation has been described by Tendolkar et al. (2004, p. 236): "[t]he first old/new effect identified during tests of recognition memory onsets approximately 400 ms post-stimulus, typically lasts around 400–600 ms, and is largest in amplitude over left temporo-parietal scalp electrodes". In our previous investigation (Campeanu et al., 2013), the voice congruency effect on memory performance coincided with the presence of this well-established ERP modulation of recollection during the word recognition test. In particular, correctly recognized old words produced a more positive deflection over parietal sites than new words, with the most positive wave occurring when old words had the voice reinstated at test. In the present investigation, then, we predict a sustained left parietal positivity for correctly recognized old words.

However, recollection is only one of two processes known to affect recognition memory. Yonelinas (1994) reviewed a dual-process model for recognition memory, in which familiarity and recollection both play a role during retrieval. Familiarity, a general strength marker that something has previously occurred, is dissociated from recollection, which usually involves retrieval of contextual details of an episode. In terms of imaging, ERP studies have typically found two distinct deflections, which were initially thought to reflect familiarity and recollection, respectively (e.g., Curran, 2000). Familiarity has been linked to a frontal positivity for old words compared to new words, between 300–500 ms (Curran, 2000; Schloerscheidt & Rugg, 2004; Curran & Dien, 2003). However, more recent research has suggested that this frontal modulation actually indexes conceptual priming rather than familiarity (Voss & Paller, 2006; Voss & Paller, 2007; Paller, Voss, & Boehm, 2007), and that the two concepts are often confounded in memory research. The present experiment does not include a behavioral measure of conceptual priming, so we cannot conclude that our participants were conceptually primed for individual words. However, it does seem reasonable to suggest that old words presented at test in the same voice as at study engage perceptual priming mechanisms. Therefore, comparing the voice congruency conditions for correctly recognized old trials presents an opportunity to investigate whether this frontal priming modulation varies as a function of perceptual congruency in a purely auditory experiment. The present investigation is the first to our knowledge to use auditory stimuli both at study and at test in an attempt to investigate this frontal deflection. If the voice congruency effect involves priming to some degree at least, we would expect to find a frontal modulation distinguishing between same-voice repetitions and different speaker trials.

Since we employ a continuous recognition paradigm, a further important objective of the present study was to investigate the effect of lag on recognition memory, and more specifically on potential voice congruency effects in word recognition. In the current experiment, we manipulated lag within each voice congruency condition, so that old words were re-presented at 2, 8 or 16 words after initial presentation. This allowed us to extend our investigation to determine how long the voice congruency benefit on word memory lasts. It is important to note that lag effects may reflect trace decay over time and/or active interference from intervening items. Regardless of the mechanism, we predict a behavioral benefit of voice reinstatement at test, as well as a decrease in performance at extended lags. In terms of the ERP trace, the amplitude of the late positive parietal (old/new) effect has been found to decrease with increasing lag (Nielsen-Bohlman & Knight, 1994). Hence, we expect electrophysiological findings to show a sustained positivity over parietal sites that decreases with increasing lag and an early frontal modulation that is sensitive to voice congruency.

2. Methods

2.1. Participants

Nineteen participants provided written informed consent according to the guidelines set out by the Baycrest Centre and the University of Toronto. EEG data from three participants were excluded because of excessive muscle artifacts and/or eye movements during recording. Data from one participant was excluded because he did not follow task instructions. The final sample of 15 participants comprised 6 males and 9 females aged between 22 and 33 (M=24.47, SD=3.38 years). All participants learned English as their first language, were right-handed and had pure-tone thresholds within normal limits for frequencies ranging from 250 to 8000 Hz (both ears), which was verified by an audiogram performed at the beginning of the session.

2.2. Stimuli

The word list consisted of 336 high-frequency (30+ from Kucera & Francis, 1967), two-syllable nouns, taken from the MRC psycholinguistic database (Wilson, 1988). All words were recorded by four speakers: one native-English female, one native-English male, one Chinese-accented female and one Chinese-accented male – in continuous streams, at 32,000 Hz sampling rate, mono, with 16-bit resolution. The native English-speaking female was 31 years old at the time of recording, and the native English-speaking male was 27 years old. The Chinese-accented female was 57 years old and had learned English at 27 years of age. The Chinese-accented male was 44 years old and had learned English at 19 years of age. With a Shure KSM44 microphone and an USBPre preamplifier with digitizer, speakers recorded the words using Adobe Audition 1.5 on a Dell laptop. All words were then spliced into individual files, using a MATLAB (The MathWorks, Inc., http://www.mathworks.com/, version 5.3) script, and their amplitudes were normalized to a standard dB level (the average of all words). The MATLAB script used to splice the words specified thresholds, duration and pre-stimulus intervals, which could be varied as necessary for different groups of words. Words were inspected for background noise, clipping, duration (1 s) and clarity. Editing was done as appropriate, either manually or automatically, using batch files. To check for intelligibility, two young adults with English as a first language each listened to three blocks of words and indicated which words were difficult to understand. Reshuffling and processing of new words occurred as necessary to ensure that potential participants – young adults with English as a first language – would have no trouble understanding the spoken words. Once all words were judged adequate, they were normalized to their average loudness, at −32.44 dB.

2.3. Experimental design and procedure

The experiment was programmed using Presentation software (Neurobehavioral Systems, http://www.neurobs.com/, version 14.5) and custom MATLAB code (version 7.1). On each trial, participants heard one word, presented binaurally through insert earphones (EAR-TONE 3a), at 70 dB SPL. Words (digitized as .wav files) were converted to analog using a computer soundcard (16 bit, stereo, with a sampling rate of 44,100 Hz). The analog output was fed into a 10 kHz filter (Tucker Davis Technologies (TDT, Alachua, FL), FT6-2), and then to a GSI 61 audiometer. The inter-stimulus interval (ISI) was jittered between 2.8 and 3.2 s (33 or 34 ms steps, rectangular distribution). Participants indicated whether or not they had previously heard the current word, by making a button press corresponding to "old" or "new" with their left or right forefinger, respectively. They were asked to judge a word as "old" if they had heard it earlier, irrespective of whether the test word was presented in the same or a different voice as its previous occurrence. They were instructed to respond as accurately as possible on each trial. As such, it is important to note that the voice manipulation was incidental in this experiment; the participant's task was simply to judge whether the current word was old or new. The experiment consisted of two 23-min blocks, each comprising 348 trials (one word per trial). No words were repeated across blocks. Participants were given a mandatory break, lasting at least two minutes, between the two blocks.