International Journal of Psychophysiology xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

International Journal of Psychophysiology

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

Speaker's voice as a memory cue

Sandra Campeanu [a,b], Fergus I.M. Craik [a,b], Claude Alain [a,b,c,*]

[a] Rotman Research Institute, Baycrest Centre, Toronto, ON, Canada
[b] Department of Psychology, University of Toronto, ON, Canada
[c] Institute of Medical Sciences, University of Toronto, ON, Canada

ARTICLE INFO

Article history:
Received 28 January 2014
Received in revised form 14 August 2014
Accepted 19 August 2014
Available online xxxx

Keywords:
Voice
Event-related potential
Memory
Context
Recollection
Attention

ABSTRACT

Speaker's voice occupies a central role as the cornerstone of auditory social interaction. Here, we review the evidence suggesting that speaker's voice constitutes an integral context cue in auditory memory. Investigation into the nature of voice representation as a memory cue is essential to understanding auditory memory and the neural correlates which underlie it. Evidence from behavioral and electrophysiological studies suggest that while specific voice reinstatement (i.e., same speaker) often appears to facilitate word memory even without attention to voice at study, the presence of a partial benefit of similar voices between study and test is less clear. In terms of explicit memory experiments utilizing unfamiliar voices, encoding methods appear to play a pivotal role. Voice congruency effects have been found when voice is specifically attended at study (i.e., when relatively shallow, perceptual encoding takes place). These behavioral findings coincide with neural indices of memory performance such as the parietal old/new recollection effect and the late right frontal effect. The former distinguishes between correctly identified old words and correctly identified new words, and reflects voice congruency only when voice is attended at study. Characterization of the latter likely depends upon voice memory, rather than word memory. There is also evidence to suggest that voice effects can be found in implicit memory paradigms. However, the presence of voice effects appears to depend greatly on the task employed. Using a word identification task, perceptual similarity between study and test conditions is, like for explicit memory tests, crucial. In addition, the type of noise employed appears to have a differential effect. While voice effects have been observed when white noise is used at both study and test, using multi-talker babble does not confer the same results. In terms of neuroimaging research modulations, characterization of an implicit memory effect reflective of voice congruency is currently lacking.

1. Introduction

Speaker's voice is well known to influence attention and memory. For instance, in noisy environments we more easily understand what a person is saying if the voice is familiar (Rosenblum et al., 2007). A familiar voice also helps us locate friends in a crowded room (Johnsrude et al., 2013). These examples highlight the impact of voice in guiding attentional processes, but the effect of voice congruency on memory is less clear. One major area of interest in memory research has been to identify factors that facilitate retrieval of information. In the context of auditory verbal memory, speaker's voice likely occupies a central role in modulating memory performance. Investigation into the nature of voice representation as a memory cue is essential to understanding auditory memory and the neural correlates which underlie it. Here, we review studies that have examined the effects of speaker's voice on memory performance in an effort to answer the following questions. Do people remember words better if they are repeated in the same voice? What about a similar voice? Does any such benefit of speaker's voice on memory performance depend on explicit attention to the voice? The present work aims to develop a recipe for finding voice effects in auditory memory studies, using neutral stimuli, in young adults, and to highlight future directions of inquiry — especially in an effort to understand the neural correlates that might underlie these voice effects in various situations. The investigation into voice – arguably the most ecologically important source in auditory scene analysis – as an auditory memory cue began some forty years ago.

To our knowledge, the study from Craik and Kirsner (1974) was among the first to examine the role of voice reinstatement on memory performance for spoken words. Using a continuous recognition paradigm, they found that participants were better and faster at recognizing words spoken in the same voice at test as at study. Though the effect was relatively small, the conclusion was that voice information persists over time and that context congruency (i.e., same speaker between study and test) facilitates verbal memory. The notion that context congruency facilitates memory performance is consistent with the encoding specificity principle (Tulving and Thomson, 1973). The principle states that the details of encoding dictate what is stored, which in turn dictates the cues that are effective in retrieval. In the case of word

* Corresponding author at: Rotman Research Institute, Baycrest Centre, Toronto, Ontario M6A 2E1, Canada. Tel.: +1 416 785 2500x3523.
E-mail address: calain@research.baycrest.org (C. Alain).

http://dx.doi.org/10.1016/j.ijpsycho.2014.08.988
0167-8760/

recognition, items are better remembered if the test episode is as close as possible to the study episode; for example, using the same speaker's voice between study and test facilitates word memory.

The findings from Craik and Kirsner (1974) raise several questions. First, does partial voice congruency provide a partial facilitation effect, or is full voice reinstatement required to see a memory benefit? The literature addressing this question has provided mixed results, as described later in the article (Palmeri et al., 1993 vs. Goldinger, 1996). Second, what circumstances (if any) modulate this facilitation effect? Third, does attention to voice matter, or are the beneficial effects of voice encoded automatically? Finally, is facilitation present for implicit memory as well, or only for explicit recognition memory?

In the following sections we review empirical findings and theoretical suggestions relevant to the use of voice as a memory cue in explicit and implicit auditory memory. With respect to explicit memory, we discuss word recognition and voice recognition and summarize neural correlates found to underlie the use of voice in such cases. With respect to implicit memory, we discuss voice-related findings found using indirect memory tests in the auditory domain. In both kinds of memory, we discuss the effects of attention allocation and expertise, in an effort to delineate conditions that promote beneficial voice effects. In conclusion, we summarize findings and suggest future directions for research investigating voice effects in memory.

2. Voice effects in explicit memory

As mentioned above, the study by Craik and Kirsner (1974) led the way for investigations into the role of voice in explicit memory. Pisoni (1993) conducted a review and concluded that detailed perceptual information about speaker voice, such as gender and dialect of the speaker, is encoded automatically and in parallel with the episodic memory trace of the word itself, and is therefore retained in long term memory. However, further research into the role of voice in word and voice memory provided mixed results.

Though they did find same-voice facilitation for word memory, Craik and Kirsner did not find a same-voice benefit for memory of the voice itself. Geiselman and Bellezza (1976, 1977) then suggested that gender of the speaker voice is automatically encoded. In their first study, Geiselman and Bellezza (1976) investigated the effect of adding an explicit voice recognition task to a sentence memory test. After free recall of sentences that had previously been heard, participants made a voice judgment for each sentence (male or female). Critically, some participants were told about the subsequent source test (intentional condition), while some were not (incidental condition). It was hypothesized that if speaker voice was encoded automatically, then participants in the incidental condition would demonstrate speaker voice recognition without a corresponding decrease in sentence recall. In fact, participants in both incidental and intentional conditions were able to recognize the voice attribute (i.e., gender) above chance levels, and did so without suffering a decrease in sentence recall performance. The authors suggested that this automaticity of voice information might be due to an integral incorporation of the voice into the code, rather than a mere attachment to the memory trace. In their subsequent study, Geiselman and Bellezza (1977) investigated whether such integration would mean that the connotation of the sentence and its relation to the gender of the speaker's voice, would make a difference. Participants heard sentences spoken by male or female voices. The sentences either contained gender-specific nouns or gender-neutral nouns. Results indicated that voice retention was superior when sentence stimuli were gender-neutral. Apparently voice is automatically encoded when neutral stimuli are employed. The authors suggested a voice connotation hypothesis — speaker voice interacts with the semantic interpretation of the sentences; automatic encoding that precludes such interaction occurs when the sentences are gender-neutral. These findings suggested that the gender of a voice is automatically encoded, and it followed that the lack of voice recognition effects in the Craik and Kirsner (1974) study might reflect a gender confound. That is, the different speaker condition in the Craik and Kirsner investigation always corresponded to a gender change, making it so that participants were just as likely to detect congruity as incongruity of the voice. Whether more detailed voice information is encoded or not, remained to be investigated.

Palmeri, Goldinger and Pisoni (1993) investigated this possibility in a follow-up to Craik and Kirsner's (1974) investigation. Since only two voices were used in the initial study by Craik and Kirsner – one of each gender – the study by Palmeri et al. (1993) sought to clarify a potential gender confound and to extend the investigation into multiple talker conditions. Using a continuous recognition paradigm, Palmeri et al. found that words repeated in the same voice during the second presentation as in the first were better recognized than words re-presented in different voices, irrespective of the gender of the voice. This finding showed that more detailed physical information, not only gender of the speaker, is encoded with the word. The investigators also manipulated the number of voices used in each continuous stream and found that talker variability did not affect performance. Interestingly, words were not better remembered if the voice at test was of the same gender as at study. Therefore, the study by Palmeri et al. indicated no partial benefit in word recognition for similar voices — at least when the partial benefit came from gender.

Palmeri et al. (1993) also concurrently measured voice recognition. Participants were asked to identify whether a test word was new, old with the same voice as at study, or old with a different voice than at study. Results indicated that a benefit in voice (or "source") memory occurred when words were presented in the same voice at study and at test. Interestingly, the study concluded that voice recognition was more accurate when words were re-presented in either the same voice or in different voices of different genders rather than in different voices of the same gender. There was, however, no difference between performance when the same voice was used and when different voices of different genders were used. This last finding echoed Craik and Kirsner's (1974) finding with respect to voice recognition. Speculatively, it may be that different voice trials can be just as perceptually salient as same voice trials, if the differences in voice characteristics are substantial (i.e., across gender).

A few years later, Dodson et al. (1998) reported evidence in support of this conjecture when they investigated memory for source information (voice) during a divided attention task. They found that memory for the voice (so-called specific source information) was compromised when attention was divided, while retrieval of the gender of the voice (so-called partial source information) remained intact. From this result, it appears that gender is a strong if general vocal cue that is automatically encoded with the memory trace, while attention plays a pivotal role in more specific voice recognition.

Therefore, an important factor in the Palmeri et al. (1993) investigation was allocation of participants' attention. Specifically, participants did not expressly attend to voice. The finding of a same speaker advantage but no partial same gender benefit might have been contingent on attention, with the exact voice reinstatement benefit in word recognition being fairly automatic while partial benefits associated with lesser voice congruency might depend on attention allocation. Indeed, and in contrast to the notion of automaticity of voice information, several studies have pointed to the importance of attention allocation for finding voice effects. In a series of experiments, Goldinger (1996) explored voice congruency as a function of levels of processing and type of memory investigated. In a first experiment, he devised a similarity matrix of voices based on participant responses — which classified voices on an ordinal scale based on pitch. This matrix was used to rate voice congruency between study and test in the subsequent experiments. Using surprise recognition tasks, Goldinger found that reinstating the same voice at test as at study improved word recognition. In addition, recognition was better when words were re-presented in a different voice of the same gender than when they were re-presented in a different voice of

Please cite this article as: Campeanu, S., et al., Speaker's voice as a memory cue, Int. J. Psychophysiol. (2014), http://dx.doi.org/10.1016/j.ijpsycho.2014.08.988