EFFECTS OF SPEAKER VARIABILITY ON PROCESSING SPOKEN WORD FORM AND MEANING IN SHORT-TERM PRIMING

Yu Zhang and Chao-Yang Lee

Ohio University, USA
yz137808@ohio.edu

ABSTRACT

Processing of spoken word form and meaning is separately evaluated from short-term repetition and semantic/associative priming experiments to investigate the role of speaker variability in spoken word recognition. The assumption that lexical representation and processing only involve abstract component devoid of stimulus variability is evaluated. The results from the repetition priming experiment show a robust attenuation of phonological form priming by speaker variability. However, the same effect is absent from the semantic priming experiment. These results suggest that the effect of speaker variability on processing spoken language may depend on the depth or level of processing. The time course for speaker variability cannot be confirmed from these two experiments. Different patterns arise from lexical decision and voice discrimination tasks, suggesting the influence of attention factors on speaker variability effect.

Keywords: speaker variability, short-term priming, voice discrimination, spoken word recognition.

1. INTRODUCTION

Spoken word recognition entails mapping acoustic signal to word entries stored in the mental lexicon. However, the nature of the entries in the lexicon has not been sufficiently elaborated to inform how human listeners store and process spoken language. Early evidence suggests that abstract representations devoid of acoustic variability mediate processing and representation of spoken words. More recent findings suggest a richer representation of spoken words and a mapping process involving preservation of stimulus variability [1]. Speaker variability, also known as indexical variability, is one type of stimulus variability embedded in the acoustic signal of speech that carries important information about individual speaker's voice quality, affective state, dialect, and more [2]. To what extent speaker variability influences real-time processing of spoken word form and meaning is an open question. In this study, we report two short-term priming experiments with lexical decision and voice discrimination tasks to shed light on this issue.

Previous research using long-term priming paradigm has suggested encoding of speaker voice information in implicit memory such that it influences performance on word recognition and identification tasks [3, 4, 5]. The magnitude of priming or repetition effect was reduced when there was a change in speaker voice between the first and second presentation of the stimulus. This line of research suggests that speaker voice may be an integrated component of spoken word form representations.

The extent to which word meaning representation is influenced by stimulus variability is lesser known. Some evidence suggests that sub-phonemic variation in voice onset time (VOT) could affect access to word meaning at inter-stimulus intervals (ISI) as short as 50 milliseconds (ms) [6]. There is also evidence that speaker variability may affect access to word meaning, although the effect was small and specific to a particular speaker's voice [7]. Does speaker variability influence word form and meaning proportionately? How does speaker variability compare with other types of variability with respect to their influence on spoken word recognition? These are the major questions addressed in this study. Given that stimulus variability may be subject to fast decay in word recognition process, investigation of the time course of speaker variability effect is also attempted in this study.

2. METHOD

Two short-term priming experiments were conducted. The first experiment was based on phonological form/repetition priming, where the items in a pair were either identical or unrelated. The second experiment was based on semantic/associative priming, where the items in a pair were either semantically associated or unrelated. The ISI was set at 50 and 250 ms to allow for a direct comparison between this study and previous findings [6, 7, 9].

2.1. Materials

The stimuli in the first experiment were adapted from Andruski, Blumstein and Burton's study [6], consisting of 26 target words paired with the same

number of prime words to generate four within-subject phonological priming conditions: (1) repeated, same speaker (e.g., “queen-queen”, “same male speaker”), (2) unrelated, same speaker (e.g., “bell-queen”, “same male speaker”), (3) repeated, different speaker (e.g., “queen-queen”, “different male speakers”), and (4) unrelated, different speaker (e.g., “bell-queen”, “different male speakers”). For the second experiment, half of the prime words were changed to formulate four within-subject semantic priming conditions: (1) related, same speaker (e.g., “king-queen”, “same male speaker”), (2) unrelated, same speaker (e.g., “bell-queen”, “same male speaker”), (3) related, different speaker (e.g., “king-queen”, “different male speakers”), and (4) unrelated, different speaker (e.g., “bell-queen”, “different male speakers”). Within each experiment, another 26 non-word targets were paired with the prime words to serve as fillers.

All stimuli were recorded by two male native speakers of American English from the same town in central Ohio. The recordings were completed in a double-walled sound-treated booth with an Audio-technica AT825 microphone, connected to a computer through a USBPre microphone interface. The average fundamental frequencies for the two male speakers were 113 Hz and 125 Hz, respectively. A t-test suggested that the two fundamental frequencies were statistically different from each other, $p < .001$. The recordings were sampled using the Brown Lab Interactive Speech System (BLISS) [8] at 22050 Hz with 14-bit quantization. Stimulus items were identified from the waveform display using MeV, the waveform editor of BLISS, and were saved as individual sound files after normalizing the peak amplitude of all the stimuli. The average duration of target words was not statistically different between the two speakers, $p = .227$.

2.2. Experiment 1: Repetition priming

The purpose of this experiment is to explore the influence of speaker variability on real-time spoken word form processing.

2.2.1. Participants

Forty college-age ($M = 20$, $SD = 2$) native speakers of American English participated in experiment 1. They passed a binaural pure tone hearing screening in a sound-treated booth at 20 dB HL across octave frequencies from 1 kHz to 4 kHz.

2.2.2. Procedure

The participants completed a lexical decision task followed by a voice discrimination task in a sound-treated booth. They were instructed in the first task to listen to pairs of items delivered through a headset and decide whether or not the second item in a pair was a real English word by pressing labelled buttons (WORD, NONWORD) on a computer keyboard. A Windows personal laptop computer equipped with the subject-testing program AVRunner in BLISS was used for stimulus delivery and response data acquisition. The participants were told to respond as quickly and accurately as possible. Prior to the actual experiment, ten practice trials were provided to familiarize the participants with the response format. To avoid possible long-term priming effects from multiple presentations of target words, four lists of 104 trials were created such that both speaker voice condition (same, different) and word relation condition (repeated, unrelated) were systematically coupled without repeating prime-target pairs in each list. Participants were randomly assigned to one of the four lists. The order of trial presentation within a list was randomized. Half of the 40 participants received the lists with ISI at 50 ms, the other half 250 ms. Immediately after the lexical decision task, the same participants were instructed to complete a voice discrimination task on the same stimuli, where they were asked to decide whether or not they hear the same speaker’s voice for the two items in a trial by pressing buttons labelled SAME or DIFFERENT on a computer keyboard. This second task was intended to focus participants’ attention on the voice rather than the linguistic information of the stimuli. The two tasks combined lasted approximately 20 minutes for each participant.

2.2.3. Results and discussion

Response accuracy and reaction time data from the two tasks were analyzed. Separate analyses of variance (ANOVAs) for the two tasks were performed on the data acquired from BLISS. Only data from real-word targets were included in the analyses. Accuracy in percentage was arcsine transformed before ANOVA. Word relation (repeated, unrelated) and speaker voice (same, different) were within-subject factors and ISI (50 ms, 250 ms) was the between-subject factor.

The response accuracy data of the lexical decision task are shown in Table 1. Overall, the lexical decision response accuracy is 97.4% (SD = 4.6). This high response accuracy indicates that participants were highly capable of distinguishing words from non-words, despite the possible

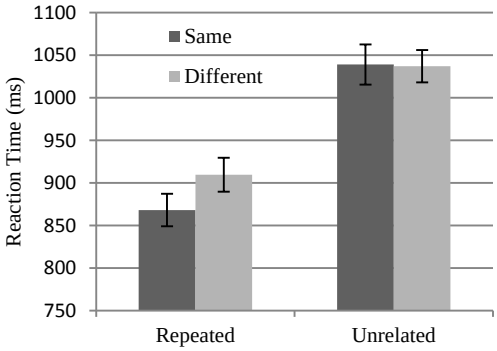
unnaturalness of the task that involves metalinguistic judgment. The main effect of word relation is significant [$F(1, 38) = 8.52$, $p = .006$, $\eta^2_p = .18$]. Participants were more accurate in their responses to repeated trials compared with unrelated trials, showing a repetition priming effect. No other effects are significant. The reaction time results of the lexical decision task are displayed in Figure 1. There is a significant main effect of word relation [$F(1, 38) = 225.795$, $p < .001$, $\eta^2_p = .86$]. The interaction between word relation and speaker voice is also significant [$F(1, 38) = 5.99$, $p = .019$, $\eta^2_p = .14$], which indicates that participants gained more facilitation from repetition trials spoken by the same speaker, relative to the facilitation from repetition trials spoken by different speakers. Moreover, the interaction between word relation and ISI is significant [$F(1, 38) = 6.325$, $p = .016$, $\eta^2_p = .14$], suggesting that there is more priming if ISI is shorter.

Table 1: Mean accuracy of lexical decision task responses (% with SD) in experiment 1.

	Same		Different	
	Repeated	Unrelated	Repeated	Unrelated
50	97 (5)	96 (6)	99 (3)	97 (6)
250	99 (3)	96 (5)	98 (4)	97 (5)
M	98 (4)	96 (5)	98 (3)	97 (5)

These results replicate previous findings that speaker variability influences processing of word form [9], with the exclusion of such confounding factors as speaker gender, since the speakers are both male in this study, and target word duration (no statistical difference for the two speakers).

Figure 1: Mean reaction time from lexical decision task as a function of word relation and speaker voice in experiment 1.



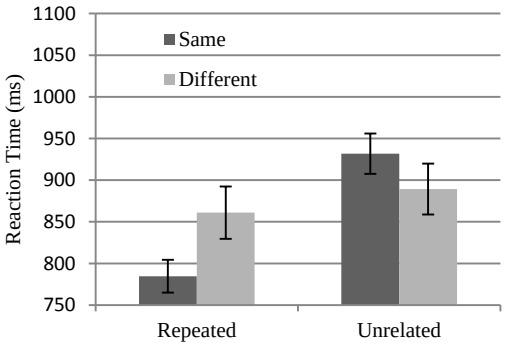
Results from the voice discrimination task are shown in Table 2 and Figure 2. Comparable to the lexical decision task, the overall response accuracy is 97.4% (*SD* = 5.7). Only the interaction between speaker voice and word relation is significant for the accuracy data [$F(1, 38) = 19.87$, $p < .001$, $\eta^2_p = .34$], suggesting that participants were more accurate for repeated trials delivered by the same voice than by a different voice. The reaction time result suggests repetition priming, as evidenced by a significant main effect of word relation [$F(1, 38) = 38.62$, $p < .001$, $\eta^2_p = .50$]. The interaction between word relation and speaker voice is also significant [$F(1, 38) = 28.29$, $p < .001$, $\eta^2_p = .43$], suggesting a disruption in priming by speaker variability. No other effects are significant.

Table 2: Mean accuracy of voice discrimination task responses (% with SD) in experiment 1.

	Same		Different	
	Repeated	Unrelated	Repeated	Unrelated
50	100 (0)	97 (5)	95 (8)	98 (3)
250	100 (2)	95 (8)	95 (8)	98 (5)
M	100 (1)	96 (7)	95 (8)	98 (4)

Results from the voice discrimination task clearly show an interaction between processing indexical variability (speaker voice) and processing of linguistic information (spoken words), even when the participants were instructed to focus only on the non-linguistic aspect of the stimuli, i.e., speaker voice.

Figure 2: Mean reaction time from voice discrimination task as a function of word relation and speaker voice in experiment 1.



2.3. Experiment 2: Semantic priming

The goal of this experiment is to explicate the effect of speaker variability on real-time processing of spoken word meaning.

2.3.1. Participants

Forty college-age (M = 22, SD = 3) native speakers of American English participated in experiment 2. None of them had participated in the repetition priming experiment. They all passed a binaural pure tone hearing screening using the same procedure as for participants in experiment 1.

2.3.2. Procedure

The procedure for this experiment is identical to experiment 1.

2.3.3. Results and discussion

Parallel to experiment 1, ANOVAs were performed on both accuracy and reaction time data collected from experiment 2. The within-subject factors were word relation (related, unrelated) and speaker voice (same, different). The between-subject factor was ISI (50 ms, 250 ms).

Only the main effect of word relation is significant from the lexical decision task, for both accuracy and reaction time data [$F(1, 38) = 11.78$, $p = .001$, $\eta^2_p = .24$; $F(1, 38) = 42.21$, $p < .001$, $\eta^2_p = .53$, respectively], indicating faster (by 53 ms on average) and more accurate (by 3% on average) responses to semantically related pairs than unrelated pairs. In contrast, the voice discrimination task yields a significant main effect of speaker voice from reaction time data only [$F(1, 38) = 14.72$, $p < .001$, $\eta^2_p = .28$], which shows faster responses (by 50 ms on average) to trials delivered by different speakers.

Results from the lexical decision task demonstrated a semantic/associative priming effect that was not affected by speaker variability. The voice discrimination result further illustrated the possible separation of voice processing and word meaning processing, in that no priming effect was observed.

3. GENERAL DISCUSSION

The purpose of this study is to explore the effect of speaker variability on spoken word recognition and lexical access. To that end, two short-term priming experiments were designed and integrated with both lexical decision and voice discrimination tasks.

The first experiment showed robust phonological form priming. The average facilitation was 149 ms. More importantly, the facilitation was attenuated by 43 ms on average from a change in speaker voice, i.e., speaker variability. This interaction still existed when the task switched to voice discrimination where attention to linguistic content was not entailed. In fact, the magnitude of priming was reduced by an average of 119 ms, although the average priming was 88 ms for the voice discrimination task.

The second experiment showed only main effects of critical variables for each task. The semantic priming averaged at 53 ms, which is smaller than the repetition priming effect obtained in the first experiment. Another important contrast is the absence of priming effect from the voice discrimination task in experiment 2. This suggests that processing of speaker variability may be unaffected by word meaning processing.

Taken together, the two experiments indicate that speaker variability influences word form and meaning processing disproportionately. The effect is influenced by levels of processing and attention factors, depending on task requirements. Since the same research assistant conducted both experiments, it is unlikely that the contrast between semantic and repetition priming results came from experimenter factors. Therefore, the more likely account may be based on the difference between phonological form level and the semantic level of spoken words. Recall that VOT variability influences access to word meaning [6], but speaker variability does not attenuate semantic priming in this study. The difference may be attributed to the fact that the VOT continuum could potentially change phonemic category membership, but speaker voice difference may be more peripheral to phoneme categorization. Because phonemes are the smallest sound units that can contrast word meaning, the differential effect can be expected. Since ISI was only interacting with repetition priming, not involving speaker variability, further research using a different ISI window is warranted to investigate the time-course hypothesis of processing speaker variability [10].

4. CONCLUSION

Based on data from two short-term priming experiments in this study, speaker variability may differentially affect word form and meaning processing. The effect is also sensitive to the manipulation of attention factors. In comparison with the stimulus variability of VOT, this study also demonstrates that not all types of variability are equally detrimental to spoken language processing.

5. REFERENCES

[1] Pisoni, D. B. 1997. Some thoughts on "normalization" in speech perception. In: Johnson K., Mullennix, J. W. (eds), *Talker Variability in Speech Processing*. San Diego, CA: Academic Press, 9–32.

[2] Abercrombie, D. 1967. *Elements of general phonetics*. Chicago: Aldine.
[3] Schacter, D. L., Church, B. A. 1992. Auditory priming: Implicit and explicit memory for words and voices. *J. of Exp. Psychol: Learn, Mem, and Cogn* 18, 915–930.
[4] Goldinger, S. D. 1996. Words and voices: Episodic traces in spoken word identification and recognition memory. *J. of Exp. Psychol: Learn, Mem, and Cogn* 22, 1166–1183.
[5] Luce, P. A., McLennan, C. T., Charles-Luce, J. 2003. Abstractness and specificity in spoken word recognition: Indexical and allophonic variability in long-term repetition priming. In: Bowers, J. S., Marsolek, C. J. (eds.), *Rethinking Implicit Memory*. New York, NY: Oxford University Press, 197–214.
[6] Andruski, J. E., Blumstein, S. E., Burton, M. 1994. The effects of subphonetic differences on lexical access. *Cognition* 52, 163–187.
[7] Zhang, Y., Lee, C.-Y. 2011. Talker variability in lexical access: Evidence from semantic priming. *Proc. 17th ICPhS* Hong Kong, 2308–2311.
[8] Mertus, J. A. 2000. The Brown Lab Interactive Speech System. Brown University.
[9] Lee, C.-Y., Zhang, Y. in press. Processing speaker variability in repetition and semantic/associative priming. *J. of Psycholing. Res*. doi: 10.1007/s10936-014-9307-5
[10] McLennan, C. T., Luce, P. A. 2005. Examining the time course of indexical specificity effects in spoken word recognition. *J. of Exp. Psychol: Learn, Mem, and Cogn*, 31, 306–321.