Proceedings of Meetings on Acoustics

Volume 19, 2013 http://acousticalsociety.org/

ICA 2013 Montreal
Montreal, Canada
2 - 7 June 2013

Speech Communication
Session 2pSCb: Speech Intelligibility (Poster Session)

2pSCb9. Continuous recognition memory for spoken words in noise

Susanne Brouwer*

***Corresponding author's address: Northwestern University, Evanston, Illinois 60201, smbrouwer@hotmail.com**

Previous research has shown that talker variability affects recognition memory for spoken words (Palmeri et al., 1993). This study examines whether additive noise is similarly retained in memory for spoken words. In a continuous recognition memory task, participants listened to a list of spoken words mixed with noise consisting of a pure tone or of high-pass filtered white noise. The noise and speech were in non-overlapping frequency bands. In Experiment 1, listeners indicated whether each spoken word in the list was OLD (heard before in the list) or NEW. Results showed that listeners were as accurate and as fast at recognizing a word as old if it was repeated with the same or different noise. In Experiment 2, listeners also indicated whether words judged as OLD were repeated with the same or with a different type of noise. Results showed that listeners benefitted from hearing words presented with the same versus different noise. These data suggest that spoken words and temporally-overlapping but spectrally non-overlapping noise are retained or reconstructed together for explicit, but not for implicit recognition memory. This indicates that the extent to which noise variability is retained seems to depend on the depth of processing.

Published by the Acoustical Society of America through the American Institute of Physics

[DOI: 10.1121/1.4798781]
Received 18 Jan 2013; published 2 Jun 2013

INTRODUCTION

Speech is inherently variable. Despite such variation, listeners are able to accurately and quickly perceive the speech signal across different speakers and situations. A key question for theories of speech perception is whether and how this surface variability is encoded in the cognitive representation of speech (see e.g. Bradlow, Nygaard, & Pisoni, 1999; Mullenix, Pisoni, & Martin, 1989; Palmeri, Goldinger, & Pisoni, 1993). For example, Palmeri et al. (1993) provided evidence that talker variability affects recognition memory for spoken words. Words repeatedly presented in the same voice were recognized as having been previously encountered more quickly and accurately than words repeated in a different voice. Similarly, listeners were also more accurate at recognizing a word as a repeated item if the word was repeated at the same versus a different speaking rate (Bradlow et al., 1999). Speech representations thus not only encode the linguistic content, but also to the so-called indexical properties or surface characteristics of speech signals such as the speaker's voice and speaking rate. This idea has been incorporated in episodic models (e.g. Goldinger, 1996; Pisoni, 1993). However, it has also been shown that speech perception is not sensitive to all sources of stimulus variation. For example, Bradlow et al. (1999) found no recognition benefit for words repeated at the same overall amplitude. Listeners only had a recognition advantage when they were asked to explicitly pay attention to changes in amplitude. This reveals that although most indexical properties are encoded and retained in memory, each property differs in the extent to which it influences recognition memory.

The aim of the current study was to examine a different type of variation that often co-occurs with speech, that is, background noise. The question is whether additive noise is encoded and represented in memory for spoken words, or if this source of instance-specificity is perceptually segregated from and not encoded in the memory representation of the spoken word. We investigated this in an implicit (Experiment 1) and in an explicit version (Experiment 2) of the continuous recognition memory paradigm.

EXPERIMENT 1

Method

Participants

Forty undergraduate students at Northwestern University participated. All listeners were native speakers of American English with no history of a speech or hearing disorder. They received $10 for their participation.

Stimulus Materials and Procedure

The stimuli consisted of 139 disyllabic spoken words of which 96 were experimental items, 15 were memory load items, 20 were filler items and 8 were practice items. Each word was uttered in isolation by a native American English speaker and the recordings were digitized at a 22050 Hz sampling rate. The duration of each word was equalized to the mean length of all words (i.e. 567 ms) and each word was low-pass filtered at 5 kHz. These marginally manipulated words were then embedded in noise consisting of a pure tone (6 kHz) or of high-pass filtered white noise (7-10 kHz). The words and noise were thus in non-overlapping frequency bands. A pretest showed that participants were able to accurately recognize these words in both types of noise (3% error rate), indicating that the items were highly intelligible.

Each word in noise was presented and repeated once (except for the fillers). This repetition occurred after a lag of 4, 8, or 16 intervening items. The repetition itself counted as the last of the intervening item. Each lag was used an equal number of times in a list. The word was repeated in the same (pure tone-pure tone or white noise-white noise) or in a different type of noise (pure tone-white noise or vice versa). The probability of a same-noise versus a different-noise repetition was equal. A list started with 15 practice trials and 30 memory load trials. Twenty filler items were randomly presented next to the 192 test trials. Listeners' task was to indicate whether each spoken word in the list was OLD (heard before in the list) or NEW (the word was new to the list). The session lasted about 15 minutes.

Analysis

A statistical analysis of the error pattern was carried out with linear mixed effects model (Baayen, Davidson, & Bates, 2008) using a logistic linking function (cf. Dixon, 2008). The fixed effects were Noise Type (tone vs. white), Repetition (same vs. different), and Lag (4, 8, 16). Participants and Items were added as random intercepts to the model.

Results and Discussion

Figure 1 shows the word recognition accuracies: percent correct OLD-Same responses for same-noise repetitions (1A) and percent correct OLD-Different responses for different-noise trials repetitions (1B) as a function of lag. Task performance was highly correct for all lags, indicating that listeners were quite easily able to recognize whether they heard a spoken word in the list before or not. The analysis showed no main effects ($p > 0.1$) and no interactions ($p > 0.1$). Listeners were thus as accurate in recognizing a word as OLD if it was repeated in the same noise as if it was repeated in different noise. These results demonstrate that variation in background noise has no effect on performance in an implicit recognition memory task. It is, however possible that speech and noise are encoded together, but that this task does not reveal this integrated encoding because the responses are dominated by access to the word instead of the noise. Since the speech and noise were in non-overlapping energy bands, participants may have paid no attention to the noise and just focused on the speech for the purposes of this task. It is thus possible that listeners could retain an instance-specific memory trace that if they are explicitly asked to do so. This is tested in Experiment 2.

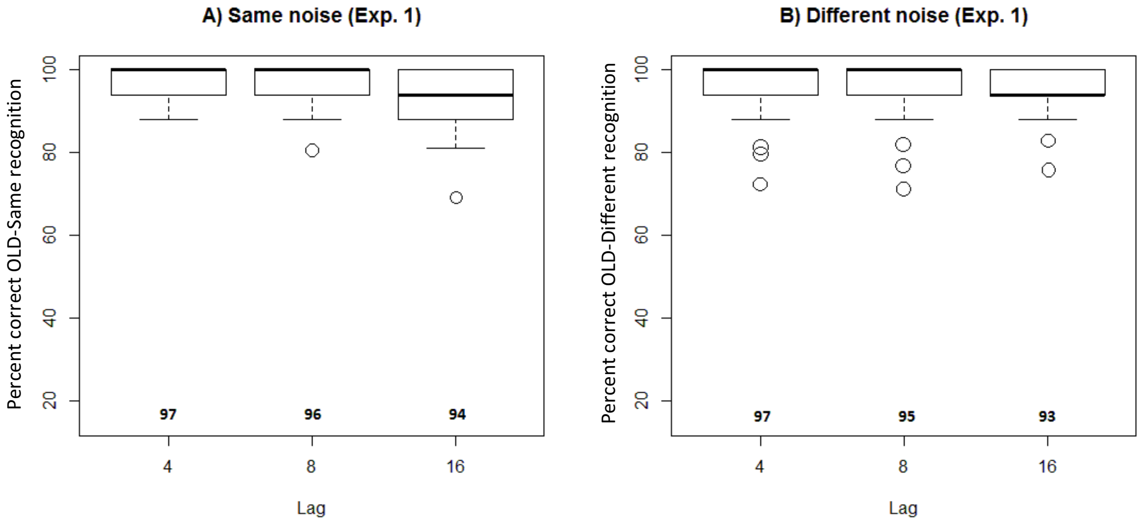

FIGURE 1. Word recognition accuracy scores as a function of lag from Experiment 1 (implicit version) for A) same-noise condition and B) different-noise condition. These boxplots show the interquartile ranges of the recognition scores. Whiskers extend to the most extreme data point that is no more than 1.5 times the interquartile range of the box. The mean is given at the bottom of each plot.

Experiment 2

Method

Participants

Twenty-four undergraduate students at Northwestern University participated. They were native speakers of American English with no history of a speech or hearing disorder and received $10 for their participation.

Stimulus Materials, Procedure, and Analysis

The stimuli were identical to the ones used in Experiment 1. As in Experiment 1, listeners indicated whether each spoken word in the list was OLD (heard before in the list) or NEW (the word was new to the list). However, for OLD responses, they were also explicitly instructed to indicate whether the word was repeated in the same or in a different type of noise. The same analysis was conducted as in Experiment 1.

Results and Discussion

Figure 2 shows the word recognition accuracies: percent correct OLD-Same responses for same-noise repetitions (2A) and percent correct OLD-Different responses for different-noise trials repetitions (2B) as a function of lag. The analysis showed a main effect of Repetition ($\beta_{\text{repetition}}= 0.28$; $p < 0.0001$), indicating that when asked to explicitly pay attention to the background noise, listeners benefitted from hearing words presented with the same versus different noise. The analysis also revealed an interaction between Repetition and Noise Type ($\beta_{\text{repetition*noisetype}}= 0.17$; $p < 0.05$) and an interaction between Repetition and Lag ($\beta_{\text{repetition*lag}}= -0.02$; $p < 0.01$). The first interaction shows a stronger effect of Repetition in the white-noise than in the tone-noise condition. The second interaction demonstrates a more robust effect of Repetition for the short (i.e. 4) than the long lag (i.e. 16). These results differ from Experiment 1 and show that same-noise repetitions increase recognition accuracy in an explicit task, especially for words in white noise and presented at short lags.

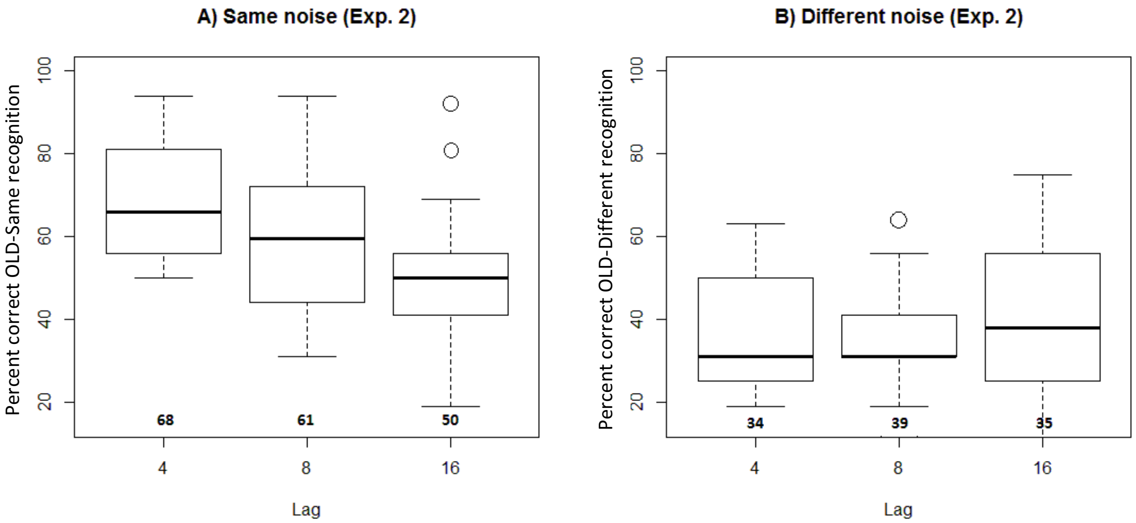

FIGURE 2. Word recognition accuracy scores as a function of lag from Experiment 2 (explicit version) for A) same-noise condition and B) different-noise condition. These boxplots show the interquartile ranges of the recognition scores. Whiskers extend to the most extreme data point that is no more than 1.5 times the interquartile range of the box. The mean is given at the bottom of each plot.

Conclusion

The aim of this study was to investigate whether listeners encode spoken words and noise together. The results from Experiment 1 show that listeners took no advantage of same-noise versus different-noise repetitions, indicating that variation in background noise has no influence on performance in a recognition memory task when the task does not require the listener to pay attention to the noise. This finding is in line with previous work that found that variation in amplitude also had no effect on recognition performance (Bradlow et al., 1999) when the listeners' attention was not drawn to amplitude variation. It could thus be the case that background noise, like amplitude, is more peripheral to speech processing than certain other sources of instance-specificity such as talker and rate variation. However, when listeners were required to pay explicit attention to the noise in the background, as in Experiment 2, there was a recognition advantage for words repeated with the same versus different background noise. Listeners' performance improved when hearing words in the same versus different noise. This suggests that spoken words and temporally-overlapping but spectrally non-overlapping noise are retained or reconstructed together for explicit, but not for implicit recognition memory. The extent to which noise variability is retained seems to depend on the depth of processing.

ACKNOWLEDGMENTS

Writing this article has been supported by Grant R01-DC005794 from NIH-NIDCD and the Hugh Knowles Center at Northwestern University. I would like to thank Ann Bradlow for useful comments on this article. I also thank Chun Liang Chan for technical support and Vanessa Dopker and Emily Kahn for help with testing participants.

REFERENCES

Baayen, R. H., Davidson, D. J., and Bates, D. M. (**2008**). Mixed-effects modeling with crossed random effects for subjects and items. Journal of Memory and Language, **59**, 390-412.

Bradlow, A. R., Nygaard, L. C., & Pisoni, D. B. **(1999)**. Effects of talker, rate, and amplitude variation on recognition memory for spoken words. Perception & Psychophysics, **61**(2), 206-219.

Dixon, P. (**2008**). Models of accuracy in repeated-measures design. Journal of Memory and Language, **59**, 447-456.

Goldinger, S. D. (**1996**). Words and voices: Episodic traces in spoken word identification and recognition memory. Journal of Experimental Psychology: Learning, Memory, and Cognition, **22**, 1166-1183.

Mullennix, J. W., Pisoni, D. B., & Martin, C. S. (**1989**). Some effects of talker variability on spoken word recognition. The Journal of the Acoustical Society of America, **85**, 365.

Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. **(1993)**. Episodic encoding of voice attributes and recognition memory for spoken words. Journal of Experimental Psychology: Learning, Memory, and Cognition, **19**, 309-328.

Pisoni, D. B. (**1993**). Long-term memory in speech perception: Some new findings on talker variability, speaking rate and perceptual learning. Speech Communication, **13**, 109-125.