THE LOCUS OF TALKER-SPECIFIC EFFECTS IN SPOKEN-WORD RECOGNITION

Alexandra Jesse [1], *James M. McQueen* [1], *Mike Page* [2]

[1]Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
[2]University of Hertfordshire, Hertfordshire, UK
alexandra.jesse@mpi.nl, james.mcqueen@mpi.nl, m.2.page@herts.ac.uk

ABSTRACT

Words repeated in the same voice are better recognized than when they are repeated in a different voice. Such findings have been taken as evidence for the storage of talker-specific lexical episodes. But results on perceptual learning suggest that talker-specific adjustments concern sublexical representations. This study thus investigates whether voice-specific repetition effects in auditory lexical decision are lexical or sublexical. The same critical set of items in Block 2 were, depending on materials in Block 1, either same-voice or different-voice word repetitions, new words comprising re-orderings of phonemes used in the same voice in Block 1, or new words with previously unused phonemes. Results show a benefit for words repeated by the same talker, and a smaller benefit for words consisting of phonemes repeated by the same talker. Talker-specific information thus appears to influence word recognition at multiple representational levels.

Keywords: Exemplar theory, Mental lexicon, Talker-specificity effects, Long-term priming.

1. INTRODUCTION

Words are more easily recognized as repetitions when they are repeated in the same voice than when they are repeated in a different voice [2]. We examine here what the source is of this talker-specific priming effect. We ask whether it reflects the storage of word-sized representations (i.e., talker-specific lexical episodes) or of sublexical representations (i.e., how talkers tend to produce individual speech sounds).

An obvious interpretation of talker-specific priming effects – in tasks such as old/new judgments on lists of spoken words or lexical decisions on lists of words and nonwords [2,3,4] – is in terms of lexical storage [2]. According to the Goldinger model [2], the surface detail of specific lexical episodes is preserved in long-term memory. This detail includes talker-specific features, but also other properties that make each token of a spoken word unique (e.g., speaking rate and contextual information). During word recognition, the incoming speech signal is compared to these detailed lexical episodes. A word is recognized better to the degree that the current signal is similar to previously-heard episodes. Since repetitions by the same talker are more similar than repetitions across talkers, this model predicts voice-specificity effects in spoken-word recognition.

Luce and Lyons [3] examined talker-specificity effects using the same materials in both an old-new recognition memory task and in lexical decision. In the memory task, listeners judged words as "old" significantly faster when the repetitions were in the same voice than when the words were in a different voice. Lexical decisions to the same words, however, were numerically but not significantly speeded for same-voice relative to different-voice repetitions. This result – that talker repetition influences memory judgments more than word recognition – could be interpreted as evidence that talker-specificity effects reflect storage in an episodic memory system that is separate from the mental lexicon. But a time-course explanation for this task difference can also be offered [4]: The high-frequency words in [3] may have been processed so quickly in lexical decision that talker-specificity effects did not have time to emerge. As evidence for this explanation, McLennan and Luce [4] showed that talker-specificity effects in lexical decision were modulated by processing difficulty (manipulated as similarity of nonword fillers to real words).

The question nevertheless remains open as to where in the cognitive system these talker-specificity effects arise. They could be due to the storage of lexical episodes in the mental lexicon [2], or to storage in a distinct (but linked) episodic memory store. Another possibility is that they reflect not the storage of word-sized representations, but rather the storage of sublexical representations.

Recent evidence from studies using a perceptual learning paradigm shows that adjustments which listeners make to talker idiosyncrasies are made at the phonemic rather than the lexical level. Listeners learned to interpret ambiguous speech sounds through disambiguating lexical knowledge [6]. For example, a fricative that was ambiguous between /s/ and /f/ ("?") was learned to be a version of /s/ when it replaced an /s/ in /s/-final words that would be nonwords if the ambiguous sound were interpreted as an /f/ (e.g., *platypu?* in English). Other listeners learned to interpret the same ambiguous sound as an /f/ when it was placed in lexical contexts favoring an /f/ interpretation (e.g., *gira?* in English).

This lexically-guided retuning of perception has two important properties. First, it can, at least under some conditions, be talker specific. An adjustment made to an ambiguous fricative based on one talker did not generalize to the processing of another talker's fricatives [1]. Second, learning transfers to new words [5]. Novel ambiguous words which could not be disambiguated lexically (e.g., *li?* in English could be either *lice* or *life*) tended to be recognized in a way that was consistent with previous training (e.g., as *lice* after /s/-biased training). This generalization of learning to new words shows that the adjustments are probably made prelexically, that is, prior to lexical access. If retuning of a phoneme in response to the training phase were to occur at the prelexical level, then that learning could automatically be applied to the recognition of all words that contain that phoneme. But if the retuning were to have a lexical or postlexical locus, then generalization of the learning to the processing of novel words would not follow automatically.

Since retuning of phoneme perception can thus be both talker-specific and prelexical, talker-specificity effects in recognition memory and lexical decision tasks may also reflect prelexical processing. The same-voice benefit in repetition priming might arise because talker-specific detail about individual phonemes is stored prelexically, such that subsequent processing of the same sounds from the same talker – which just happen to be in the same words – is facilitated.

The goal of the present study was therefore to examine the locus and nature of talker-specificity effects in lexical decision. The same critical set of test words were preceded, across four groups of participants, by either the same words spoken by the same talker, the same words spoken by another talker, different words using the same phonemes as in the test set and spoken by the test talker, or different words with different phonemes. If talker-specific knowledge influences lexical access, lexical decisions to same-voice repetitions should be faster than to different-voice repetitions. If talker-specific knowledge is stored only in word-sized representations (either in the lexicon or separately), then lexical decisions should be equally fast in the two conditions where the test items were new words, that is, independently of whether or not those words consist of previously heard phonemes. But if knowledge about how talkers produce individual phonemes is stored prelexically, then lexical decisions to the test words should be faster when they contain phonemes that have been heard before. That is, just as in [5], talker-specific sublexical knowledge should generalize to new words and lead to a recognition advantage for those words.

2. METHOD

2.1. Subjects

128 native Dutch speakers were paid to participate.

2.2. Stimuli

A test set of ten Dutch CVC words was selected so that all words contained only the phonemes {b, k, m, p, t, l, r, n, ɑ, ɔ, ɪ, aː}. A second set of ten words contained re-orderings of these twelve phonemes. The phonemes in this second set occurred exactly the same number of times as in the test set. Note also that, across the two sets, each consonant occurred equally often in onset and coda position. Another set of five words and five nonwords containing only these critical phonemes were chosen. In addition, twenty-five words and twenty-five nonwords that consisted only of phonemes that were not in the critical set were selected. All items had a CVC structure. Nonwords differed by one phoneme from real Dutch words and were phonotactically legal in Dutch. All stimuli were recorded by a male and by a female native Dutch speaker.

2.3. Design

Type of repetition was a between-subject factor. The experiment consisted of two blocks. Each block contained an equal number of words and nonwords, half of each spoken by each talker. All participants were presented with the same items in Block 2, among them the test set of ten words, and with the same items in Block 1, except for one set of ten words that differed across four groups.

For one group (*same words, same voice*), their Block 1 set was identical to the test set in Block 2. The test words were therefore exact repetitions. The second group (*same words, different voice*) was presented with the same words in their Block 1 set, but spoken by a different talker. This group had heard the test words before, but not spoken by the same talker. This group also had not previously heard any critical phonemes by the test talker. The third group (*same phonemes, same voice*) heard in their Block 1 set ten different words that consisted of re-orderings of the phonemes used in test words. This group had never heard the test words before, but had heard the same phonemes spoken by the same talker. This group heard each of the phonemes in the test words spoken equally often by this talker during the first block as had the group in the *same words, same voice* condition. The fourth, control group heard in their Block 1 set neither the same words nor the same phonemes as were used in the test words. For this group, the test set thus consisted of new words containing previously unused phonemes.

All four of these conditions were tested twice, once with the female speaker and once with the male speaker as the talker of the critical test set. Participants were assigned pseudo-randomly to one of the resulting eight groups. The presentation of trials within each block was randomized for each participant. Participants perceived the experiment as consisting of only one block.

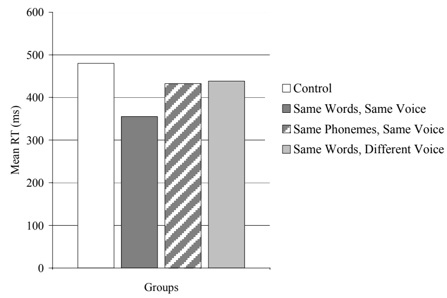
2.4. Procedure

Participants were instructed to listen to a list of words and nonwords presented over headphones at a comfortable listening level. The task was to indicate on each trial as fast and as accurately as possible if the stimulus was a Dutch word. Answers were given by pressing keys labeled ‘yes’ and ‘no’ on a button box; ‘yes’ responses were given with the participant’s dominant hand. Trial procedure was identical to that in [3]. Each trial started with the display of a fixation cross for 500 ms to signal the beginning of a trial. This was followed by a black screen, shown for 250 ms, and then the auditory stimulus. Participants had 2,200 ms to respond; otherwise a missing response was recorded. If a response was given, the next trial began after 500 ms.

3. RESULTS

Figure 1 shows mean RTs for correct ‘yes’ responses to the set of ten test words in each of the four conditions, pooled over talker. RTs were measured from the offset of the audio files. ANOVAs with repetition condition as between-subject and within-item factors, and subjects and items as random factors, respectively, were conducted. Error rates were not analyzed (<4.4%).

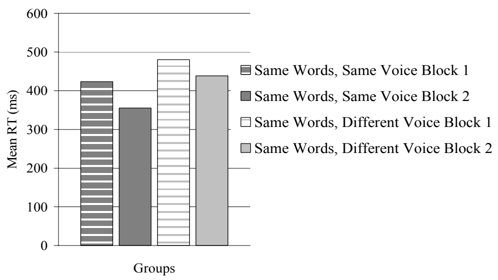
Figure 1: Lexical decision latencies across conditions.



There were main effects of repetition condition ($F_s(3,120)=5.71$, $p<.001$; $F_I(9,27)=38.61$, $p<.001$) and talker ($F_s(1,120)=15.27$, $p<.001$; $F_I(1,9)=3.60$, $p=.09$). There was also an interaction between these two factors ($F_s(3,120)=3.43$, $p<.05$; $F_I(3,27)=13.61$, $p<.001$). Planned comparisons between the three experimental conditions and the control showed that words repeated in the same voice were recognized faster than when presented for the first time ($t_s(62)=3.43$, $p<.001$; $t_I(9)=14.82$, $p<.001$). The advantage of hearing the same words but by a different talker was less robust ($t_s(62)=1.16$, $p=.13$; $t_I(9)=2.63$, $p<.05$). But there was an additional voice advantage for words repeated in the same rather than a different voice ($t_s(62)=2.62$, $p<.01$; $t_I(9)=6.73$, $p<.001$). Importantly, the repetition of phonemes also led to a recognition advantage ($t_s(62)=1.36$, $p=.09$; $t_I(9)=5.18$, $p<.001$). The benefit from words repeated in the same voice, however, was larger than for phonemes repeated in the same voice ($t_s(62)=2.49$, $p<.01$; $t_I(9)=8.67$, $p<.001$).

In addition, within-subject priming effects across blocks were analyzed for the *same words, same voice* and the *same words, different voice* groups (see Figure 2). A within-subject priming effect for the *same phonemes, same voice* group could not be analyzed because, due to the nature of the manipulation, different words were used in the two blocks. Results showed a significant difference between the two conditions ($F_s(1,18)=6.41$, $p<.05$; $F_I(1,31)=4.81$, $p<.05$). The main effect of priming was also significant ($F_s(1,18)=25.67$, $p<.001$, $F_I(1,31)=23.23$, $p<.001$), but this effect did not vary across conditions ($F_s(1,18)=1.69$, $p=.21$; $F_I(1,31)=1.82$, $p=.19$). That is, repetition of words speeds up lexical decisions no matter whether words are repeated in the same or a different voice. This result is different from the voice-specificity result found in the between-subject comparison.

Figure 2: Repetition effects in lexical decision latency for the two *same words* conditions across blocks.



4. DISCUSSION

This study provides further evidence on talker-specificity effects in the recognition of words. It presents in addition a novel result: that same-voice repetition effects are due, at least in part, to repetition of the components of words. These findings have two primary implications, one methodological, and one theoretical.

The methodological implication concerns the difference found between the between-subject analysis and the within-subject analysis. As in previous studies [3,4], lexical decisions to repeated words were faster for same-talker than for different-talker repetitions. This effect emerged in the between-subject analysis comparing the *same words, same voice* and *same words, different voice* groups. Within these two groups, however, responses to the test words sped up by the same amount whether the words were repeated by the same or a different talker. Ideally, the outcomes of the analyses would have agreed. Given that they do not, it is impossible to draw any strong conclusions about the robustness of the same-speaker repetition effect. But this conflict does provide one reason why previous studies with a within-subject design may have failed to find reliable talker-specificity effects [3].

The phoneme-repetition benefit has an important theoretical implication. This is that talker-specificity effects cannot be explained through the claim that storage of talker-specific information is based on word-sized representations alone. A model with no sublexical representations cannot account for these data. But the phoneme-repetition benefit was smaller than that when the same words repeated. A model in which the only kind of stored talker-specific knowledge is that concerning individual phonemes therefore also cannot account for these data.

Several models could account for this additional effect of same-word repetition. One possibility is a model that would assume that episodes with talker-specific detail are stored not only prelexically but also in the lexicon. Another way to account for this pattern of results would be to assume that talker details are stored in an episodic memory system that is separate from but linked to a linguistically abstract lexicon, such that those details can still influence word recognition. A third possibility is that the prelexical level has access to talker-specific knowledge both about individual speech sounds and about how different talkers put sounds together (e.g., idiosyncratic forms of coarticulation). Future research (preferably using within-subject designs) will be required to distinguish among these alternatives. The present results nevertheless already show that models of spoken-word recognition need to account for listener sensitivity to talker-specific detail at the level of individual speech sounds.

5. REFERENCES

[1] Eisner, F., McQueen, J.M. 2005. The specificity of perceptual learning in speech processing. *Perc. & Psychophys.* 67, 224-238.

[2] Goldinger, S.D. 1998. Echoes of echoes? An episodic theory of lexical access. *Psych. Rev.* 105, 251–279.

[3] Luce, P.A., Lyons, E.A. 1998. Specificity of memory representations for spoken words. *Mem. & Cog.* 26, 708-715.

[4] McLennan, C.T., Luce, P.A. 2005. Examining the time course of indexical specificity effects in spoken word recognition. *J. Exp. Psych.: Learn., Mem., Cog.* 31, 306-321.

[5] McQueen, J.M., Norris, D., Cutler, A. 2006. Phonological abstraction in the mental lexicon. *Cog. Sci.* 30, 1113-1126.

[6] Norris, D., McQueen, J.M., Cutler, A. 2003. Perceptual learning in speech. *Cog. Psych.* 47, 204-238.