Atten Percept Psychophys
DOI 10.3758/s13414-015-0854-0

Attention modulates specificity effects in spoken word recognition: Challenges to the time-course hypothesis

Rachel M. Theodore · Sheila E. Blumstein · Sahil Luthra

Abstract Findings in the domain of spoken word recognition have indicated that lexical representations contain both abstract and episodic information. It has been proposed that processing time determines when each source of information is recruited, with increased processing time being required to access lower-frequency episodic instantiations. The time-course hypothesis of specificity effects has thus identified a strong role for retrieval mechanisms mediating the use of abstract versus episodic information. Here we conducted three recognition memory experiments to examine whether the findings previously attributed to retrieval mechanisms might instead reflect attention during encoding. The results from Experiment 1 showed that talker-specificity effects emerged when subjects attended to the individual speakers, but not when they attended to lexical characteristics, during encoding, even though processing times at retrieval were equivalent. The results from Experiment 2 showed that talker-specificity effects emerged when listeners attended to talker gender but not when they attended to syntactic characteristics, even though the processing times at retrieval were significantly longer in the latter condition. The results from Experiment 3 showed no talker-specificity effects when all listeners attended to lexical characteristics, even when processing at retrieval was slowed by the addition of background noise. Collectively, these results suggest that when processing time during retrieval is decoupled from encoding factors, it fails to predict the emergence of talker-specificity effects. Rather, attention during encoding appears to be the putative variable.

R. M. Theodore (✉)
Department of Speech, Language, and Hearing Sciences,
University of Connecticut, Unit 1085, Storrs, CT 06269-1085, USA
e-mail: rachel.theodore@uconn.edu

S. E. Blumstein · S. Luthra
Department of Cognitive, Linguistic, and Psychological Sciences,
Brown University, Providence, RI, USA

Published online: 01 April 2015

Keywords Speech perception · Spoken word recognition · Talker-specificity · Attention

One pervasive theme across psychological domains concerns the cognitive factors that underlie the perceptual ability to treat physically distinct elements as members of the same conceptual category. Within the domain of spoken word recognition, a primary target of research has been to describe how listeners achieve stable perception, given the marked variability in mapping between the speech signal and linguistic representations. The acoustic–phonetic information used to specify a particular consonant or vowel, and thus for individual words, can vary from utterance to utterance depending on many factors, including speaking rate (Miller, 1981), phonetic context (Delattre, Liberman, & Cooper, 1955), and even idiosyncratic differences in pronunciation across individual talkers (e.g., Klatt, 1986; Peterson & Barney, 1952; Theodore, Miller, & DeSteno, 2009). Given this variability, the challenge for the listener is to recognize physically distinct objects as being equivalent, in order to achieve robust perception.

The prevailing theoretical view for many years was that perceptual constancy for spoken language was achieved via a normalization process, such that variability in the speech signal was discarded early in the perceptual process in order to map the speech signal onto abstract linguistic representations (e.g., Ladefoged & Broadbent, 1957; Magnuson & Nusbaum, 2007; Mullennix, Pisoni, & Martin, 1989). Under such an account, information about the specific phonetic details of an utterance was thought to be absent from long-term memory. However, more recent investigations have suggested that listeners do retain surface characteristics for individual words (Goldinger, 1998; Palmeri, Goldinger, & Pisoni, 1993), which supports episodic-based models that have posited that fine-grained phonetic information is retained in

memory (e.g., Goldinger, 1996, 1998; Grossberg, 1986). The common characteristic of these models is that each presentation of a given word is stored as a trace in memory; over time, lexical representations are viewed as a distribution centered on the most frequent experience, while also retaining specific characteristics of infrequent traces.

In this vein, a series of studies has focused on listener sensitivity to phonetic variation associated with individual speakers. It has long been known that familiarity with talkers' voices benefits subsequent processing. Not only is word intelligibility improved for familiar as compared to unfamiliar voices (Nygaard, Sommers, & Pisoni, 1994), but processing time is faster for familiar than for unfamiliar voices (Clarke & Garrett, 2004). These effects have been explained as the consequence of encoding talker-specific phonetic detail, and indeed, there is strong evidence that many detailed surface characteristics, including those associated with individual talkers, are preserved in memory (e.g., Church & Schacter, 1994; McLennan & Luce, 2005; Nygaard, Burt, & Queen, 2000; Palmeri et al., 1993; Schacter & Church, 1992).

Recent findings have suggested that such talker-specificity effects, though robust, arise relatively late in processing. Using a long-term repetition-priming paradigm, McLennan and Luce (2005) found that talker-specificity effects were observed only when processing was relatively slow. In contrast, allophonic-specificity effects were observed when processing was relatively fast (McLennan, Luce, & Charles-Luce, 2003). McLennan and colleagues explained this difference in terms of the relative frequencies of both types of variability. They posited that allophonic variability, such as a flap produced for medial /t/, is more frequently encountered than any particular talker's phonetic signature. They modeled this effect using the architecture of adaptive resonance theory (ART; Grossberg, 1986). Within the ART framework, more-frequent representations will spread activation with greater intensity, thus building to a threshold of response in advance of less frequent representations. Additional support for the time-course hypothesis has come from Mattys and Liss (2008), who manipulated processing time in a recognition memory experiment by presenting normal speech to one group of listeners and impaired speech to a different group of listeners. Reaction times were longer for the impaired than for the normal speech, and only listeners who heard impaired speech demonstrated a talker-specificity effect in recognition memory. More recently, the time-course hypothesis has been evaluated in the context of native and foreign-accented speech. The results from a lexical decision task showed a talker-specificity effect for foreign-accented speech but not for native speech, concomitant with slower processing times for the foreign-accented than for the native speech (McLennan & González, 2012).

In its current form, the time-course hypothesis of McLennan and colleagues posits that the relationship between abstract and episodic information is specified by frequency, such that the abstract source of information is always more frequent than a particular episodic trace. Accordingly, the time-course hypothesis predicts that if a response is elicited relatively early in the processing stream, abstract information will prevail, but if a response is elicited relatively late in the processing stream, the lower-frequency, episodic information will prevail, and performance will show specificity effects.

In their initial examination of the time-course hypothesis, McLennan and Luce (2005) used task difficulty to manipulate processing time, with an easy task being used to generate "fast" processing times and a difficult task to generate relatively slower processing times. For example, listeners completed a lexical decision task in which the nonwords either were very similar to real words, and thus were difficult to identify as nonwords, or were maximally distinct from real words, and thus were easily identified as nonwords. Other tasks used to manipulate processing time have included immediate versus delayed shadowing, in which the task difficulty was relatively increased in the delayed shadowing condition due to increased demands on working memory. Using task difficulty to manipulate processing time has continued in recent examinations of the time-course hypothesis (e.g., Krestar & McLennan, 2013). In the memory literature, task difficulty has been associated with encoding mechanisms such as depth of processing (Craik & Tulving, 1975). This raises the possibility that the specificity effects that emerged with slow processing times during retrieval may have been a consequence of encoding factors, and not of processing time per se.

Rather than explicitly manipulating processing time through task difficulty, Mattys and Liss (2008) manipulated it by varying the nature of the stimuli presented for the fast versus slow conditions. The stimuli in the fast condition were typical speech, and the stimuli in the slow condition were dysarthric speech, which may have been an implicit manipulation of task difficulty. Indeed, those who heard dysarthric speech had a much lower hit rate than did those who heard typical speech, suggesting that processing dysarthric speech was much more difficult than processing typical speech. The specificity effect for the dysarthric speech was observed even when only those items that were correctly identified in intelligibility pretests were analyzed, which indicates that the effect was not solely driven by intelligibility. Indeed, additional analyses showed that the specificity effect for the dysarthric speech was limited to the slow responders and was not observed for the participants with the fastest response latencies. However, the slow responders in the typical-speech condition did not show a specificity effect, which raises the possibility that the degraded signal presented with dysarthric speech may have implicitly increased attention or cognitive effort during encoding. Another example of using stimulus variation to manipulate processing time came from McLennan and González (2012), who examined the processing of native and foreign-

accented speech. Their experiments used the "easy" (and thus, "fast") lexical decision task of McLennan and Luce (2005). The critical manipulation was that one group of listeners was presented with items produced by a native speaker and the other group was presented with items produced by a nonnative speaker. Talker-specificity effects emerged only for the accented speech, concomitant with increased processing time relative to the listeners who heard native speech. Given the literature demonstrating increased difficulty in processing foreign-accented relative to native speech (e.g., Munro 1998; Munro & Derwing, 1995), it is possible that there were signal-driven differences in task difficulty between the two listener groups, despite holding the task constant. Linking task difficulty with processing time is problematic, in that it leads to difficulties in interpreting the causal relationship between time of processing and the source of information used to guide a particular response. Hence, in the present work we sought to evaluate the time-course hypothesis in a case in which processing time was decoupled from task difficulty.

To date, the literature on the time-course hypothesis has focused on the processing time at retrieval. However, a large body of evidence has indicated that the observable behavior in a memory task reflects not only retrieval mechanisms, such as processing time, but may also reflect memory encoding mechanisms. It is possible, then, that the previous findings attributed to differences in processing times during retrieval may actually have been the consequences of differences in encoding factors. In the present work, we tested this hypothesis. We used the recognition memory paradigm of Mattys and Liss (2008) to examine the role of attention during encoding in the subsequent emergence of specificity effects during lexical retrieval. Three experiments were conducted, each consisting of an encoding phase and a recognition phase. The stimulus set consisted of words produced by two healthy, native English speakers and was held constant across the three experiments. In Experiments 1 and 2, we manipulated attention during encoding, such that one group of listeners attended to talker gender and the other group attended to either lexical (Exp. 1) or syntactic (Exp. 2) aspects of the signal. Following encoding, all participants completed a recognition memory test in which they were asked to indicate on each trial whether they had heard that word during encoding. In Experiment 3, attention during encoding was directed toward lexical characteristics for two groups of listeners. After the encoding phase, half of the listeners completed the recognition task in quiet, and the other half completed the recognition task in background noise. In all three experiments, we measured the degree to which the hit rates and reaction times at recognition were influenced by whether voice was held constant for a given word between the encoding and recognition phases. If, as is predicted by the time-course hypothesis, specificity effects associated with the use of episodic information are determined by the processing time during retrieval, then we should only observe a specificity effect for listeners who completed the recognition task in background noise, and thus had the slowest processing times. If, however, specificity effects reflect the role of attention during encoding, then we would observe specificity effects only when listeners had attended to talker identity, and they would emerge irrespective of processing time and be observed even when processing was relatively fast.

Experiment 1

Two groups of listeners participated in a recognition memory task that consisted of an encoding phase and a recognition phase. The recognition phases were identical for both groups of listeners and replicated the "fast" condition used in Mattys and Liss (2008). Across the two groups of listeners, attention was manipulated during the encoding phase by directing one group to attend to individual words and the other group to attend to the talkers who were producing them. Thus, Experiment 1 was designed to test the time-course hypothesis in a case in which attention during encoding was manipulated orthogonally to the processing time during retrieval and, critically, to do so in a "fast" condition. Because the recognition phases were identical for both encoding groups, we predicted that reaction times (RTs) during recognition would not differ between the two groups of listeners. Thus, according to the time-course hypothesis, specificity effects should fail to emerge for both groups of listeners, given that their responses were elicited early in the processing stream in a "fast" condition. If, however, attention during encoding influences subsequent recognition memory, then we predicted that talker-specificity effects would be observed only for listeners who had attended to the talker characteristics during encoding, despite their having equivalent (and fast) RTs, as compared to listeners who had attended to general lexical characteristics.

Method

Subjects Twenty-four subjects were recruited from the Brown University community. Half were assigned to the lexical encoding condition, and the other half were assigned to the talker identification encoding condition. All listeners were right-handed, monolingual native speakers of American English with no history of speech, language, or neurological disorder. An additional two listeners participated but were excluded from the analyses because they did not meet the criterion for recognition hit rate, as described below.

Stimuli The stimuli included 40 monosyllabic words with a consonant–vowel–consonant syllable structure and are listed in the Appendix. The words were selected to be familiar, to

exhibit a range of phonological variation, and to share minimal semantic relatedness. Two talkers, a male and a female, were recorded producing three repetitions of each word. The talkers were native speakers of American English and had perceptually distinct voices. Speech was recorded via microphone (Sony ECM-MS907) onto a high-definition digital recorder (Roland Edirol R-09HR) and transferred to computer for analysis. The Praat speech-processing software (Boersma & Weenink, 2011) was used to isolate each word, and the best repetition of each word for each talker was selected. For the selected words, the mean fundamental frequency for the female talker was 185 Hz (*SD* = 28), and the mean fundamental frequency for the male talker was 114 Hz (*SD* = 19). The mean word duration for the female talker was 474 ms (*SD* = 66), and the mean word duration for the male talker was 424 ms (*SD* = 52).

Design Two blocks of 30 stimuli were presented, one during the encoding phase and one during the recognition phase. The blocks were constructed such that during the recognition phase, 20 words had previously been presented during encoding ("old" words), and ten words had not ("new" words). For the "old" words, voice was held constant between encoding and recognition on half of the trials (*same-talker* trials; e.g., dog_{male} during encoding and dog_{male} during recognition), and voice differed across the two phases for the other half (*different-talker* trials; e.g., dog_{male} during encoding and dog_{female} during recognition). For the "new" words, different lexical items were presented as encoding–recognition pairs, with voice being held constant for both words (e.g., dog_{male} during encoding and gas_{male} during recognition). There were equal numbers of same-talker and unrelated trials for each of the two voices. For the different-talker trials, half consisted of a particular word presented in the male voice during encoding and the female voice during recognition, and the other half followed the opposite pattern of presentation. Accordingly, each of the encoding and recognition phases consisted of equal numbers of items produced by each of the two talkers. The 40 lexical items used in this experiment were randomly assigned to a particular trial type (e.g., same-talker trial) separately for each subject, so that a given subject only heard a given word for a particular trial type. Following this assignment, the order of presentation of items for the encoding and recognition phases was randomized for each subject, with the constraint that the first item in the recognition phase was a "new" word.

Procedure All listeners were tested individually in a sound-attenuated booth and were seated in front of a response box. The auditory stimuli were presented binaurally via headphones (Sony MDR-V6) at a comfortable listening level that was held constant across subjects (59 dB SPL). All of the subjects completed an encoding phase followed by a recognition phase. The listeners in the lexical encoding condition were instructed to listen carefully to each word and to press a button to advance to the next word. The listeners in the talker identification encoding condition were instructed to listen carefully to each word and to indicate the gender of the talker by pressing the appropriately labeled button on the response box. The recognition phases were identical for listeners in both encoding conditions; all were directed to indicate on each trial whether or not the word had been presented during encoding by pressing a button labeled "yes" or "no." The button assignments were adjusted for each participant, such that the dominant hand was always used for "yes" responses. Listeners were told to ignore voice differences between encoding and recognition in making their decision and to indicate their response as quickly as possible without sacrificing accuracy. For both the encoding and recognition phases, the pause between trials was 2000 ms, timed from the button response. A very short break (approximately 2–3 min) was interposed between the two phases.

Results

Hit rate We analyzed the performance during the recognition phase for both groups of listeners as follows. The mean hit rates were calculated for each subject for same-talker and different-talker trials. We required performance during recognition to be above chance, setting the criterion for inclusion as hit rates greater than .60 for both same-talker and different-talker trials. Two of the subjects were replaced because they failed to meet this criterion.

Figure 1 (left panel) shows the mean hit rates across listeners for same-talker and different-talker trials separately for each encoding condition. The mean hit rates were submitted to an analysis of variance (ANOVA) with the between-subjects factor Encoding Condition (lexical, talker identification) and the within-subjects factor Trial Type (same talker, different talker). The results of the ANOVA showed no main effect of trial type [$F(1, 22) = 3.09$, $p = .093$, $\eta^2 = .100$] and, critically, no main effect of condition [$F(1, 22) = 0.35$, $p = .562$, $\eta^2 = .017$], the latter result indicating that directing listeners to attend to the word or to the talker did not influence overall recognition memory. However, we found a significant interaction between condition and trial type [$F(1, 22) = 6.056$, $p = .022$, $\eta^2 = .195$]. Planned comparisons were conducted in order to determine that nature of the interaction. Here, and throughout all experiments, we applied the Bonferroni correction for multiple comparisons ($\alpha = .025$). The results showed that the interaction was due to the hit rate for same-talker trials being significantly higher than that for different-talker trials in the talker identification encoding condition [.88 vs. .78, respectively; $t(11) = 2.71$, $p = .020$, $d = 0.989$], but not in the lexical

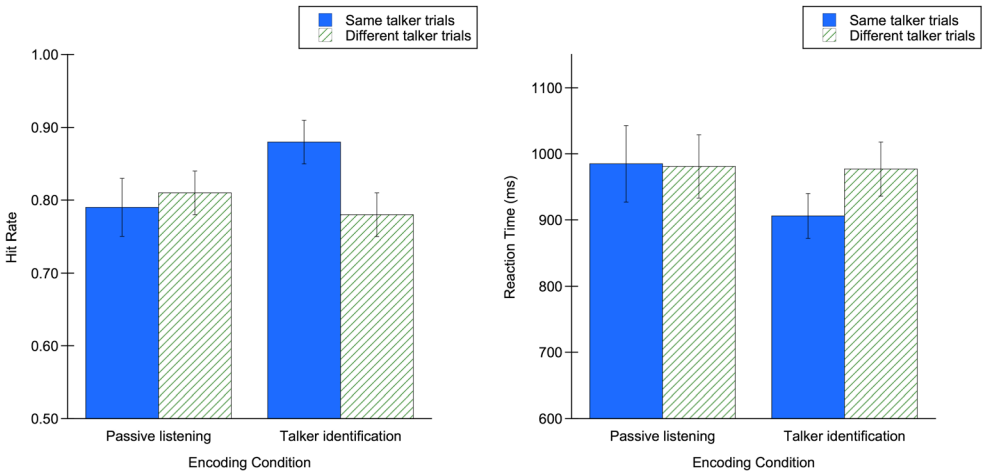

Fig. 1 Mean hit rates (left panel) and reaction times (in milliseconds, right panel) for hits during the recognition phase of Experiment 1, for each encoding condition and for same-talker and different-talker trials. Error bars indicate standard errors of the means

encoding condition [.79 vs. .81, respectively; $t(11) = 0.56$, $p = .586$, $d = -0.149$].[1]

Reaction time The RT for each trial was measured as the time between the onset of the auditory stimulus and the onset of the button response. For each subject, RTs greater than two standard deviations above the mean RT for same-talker and, separately, for different-talker trials were considered outliers and removed from subsequent analysis. Sixteen data points (4.1 % of the total data) were removed for this reason. The right panel of Fig. 1 shows the mean RTs to same-talker and different-talker trials for each encoding condition. The data were submitted to an ANOVA with Encoding Condition as a between-subjects factor and Trial Type as a within-subjects factor. The ANOVA showed no main effect of encoding condition, indicating that the RTs were equivalent across the two listener groups [$F(1, 22) = 0.43$, $p = .518$, $\eta^2 = .019$]. There was a marginal main effect of trial type [$F(1, 22) = 3.76$, $p = .066$, $\eta^2 = .124$] and a significant interaction between trial type and encoding condition [$F(1, 22) = 4.615$, $p = .043$, $\eta^2 = .152$]. Planned comparisons showed that the interaction was due to faster RTs to same-talker than to different-talker trials in the talker identification condition [906 vs. 977 ms, respectively; $t(11) = 2.74$, $p = .019$, $d = -0.0543$], but not in the lexical condition [985 vs. 981 ms, respectively; $t(11) = 0.16$, $p = .877$, $d = 0.021$].

[1] Here and throughout, prior to conducting the ANOVA comparing performance between the two listener conditions, we first conducted an ANOVA for each listener condition in order to examine whether performance in the experiment differed as a function of the two talkers' voices. For these analyses, the mean hit rate and mean RT were submitted to a repeated measures ANOVA with the factors Talker (male, female) and Trial Type (same talker, different talker). In no case did the ANOVA reveal a main effect of talker or an interaction between talker and trial type ($ps > .10$ in all cases). Accordingly, we collapsed across talkers in order to perform the analyses presented in the main text.

Discussion

When attention was explicitly directed toward talker characteristics during the encoding phase, listeners demonstrated a processing advantage for the recognition of words that had been presented in the same voice between encoding and recognition, as compared to when the voices differed across the two phases. This specificity effect indicates that listeners relied on specific episodic representations to facilitate lexical recognition. In contrast, the listeners who attended to more general lexical characteristics during encoding did not show a specificity effect during recognition. The processing times at recognition for both groups of listeners were equivalent. These data are not consistent with the predictions of the time-course hypothesis, in that a specificity effect emerged for the talker identification group in the absence of a delay in processing time relative to the lexical group. Given that overall processing times did not differ between the two listener groups, as measured by the RTs during recognition, these findings suggest that attention, and not processing time, drove the presence or absence of the specificity effect, and hence determined the use of abstract versus episodic information during the recognition task.

An alternative explanation is that attention to talker identity per se was not what gave rise to the specificity effects, but rather they arose from the consequences of requiring participants to make a decision during encoding that led to specificity effects at recognition. Recall that the listeners in the talker identification group were required to make a talker gender decision on every trial during encoding; in contrast, the listeners in the lexical encoding condition were directed to listen to each word and to press a button to advance to the next trial. This could have potentially led to a situation in which those in the talker identification encoding condition were forced to attend to the stimuli overall in order to make a decision on every trial, whereas those in the lexical encoding condition were not actually attending to lexical characteristics, as we

had intended, but were simply pressing a button to move to the next trial. To address this possibility, we analyzed the RTs during encoding (measured from the onset of the auditory stimulus to the onset of the buttonpress) and found that the processing time was significantly longer in the lexical than in the talker identification encoding condition [1471 vs. 957 ms, respectively; $t(22) = 2.49$, $p = .021$, $d = 1.015$]. This finding suggests that participants in the lexical encoding condition were indeed listening and attending to the stimuli, and not simply pressing the button to advance to the next trial as quickly as possible. However, these results do not rule out the possibility that requiring a decision in the talker identification encoding condition was responsible for the specificity effect at recognition. In Experiment 2, we directly examined this possibility.

Experiment 2

The results from Experiment 1 were not in line with the predictions of the time-course hypothesis. Specifically, a talker-specificity effect was observed in a generic "fast" condition when attention was directed toward talker identity, but not when attention was directed toward general lexical characteristics. In order to ensure that this pattern of results was not due to differences in the task demands of the two conditions in Experiment 1 (i.e., only requiring a decision to be made in the talker identification encoding condition), in Experiment 2 we examined the role of attention in a case in which the listeners were always required to make a decision during encoding.

Two groups of listeners participated in encoding and recognition phases similar in design to those in Experiment 1. One group of listeners was required to make a syntactic decision during encoding, and the other group was required to make a talker decision during encoding. Following this phase, all listeners participated in identical recognition memory tasks, as we described for Experiment 1. If, as is suggested by the results of Experiment 1, attention during encoding influences the emergence of specificity effects at recognition, then we predicted that a talker-specificity effect would only be observed for those who attended to talker identity. If however, the results from Experiment 1 reflected only the consequences of making a judgment during encoding, irrespective of attention demands, then we predicted that talker-specificity effects would emerge for both groups of listeners.

Method

Subjects Twenty-four subjects who had not participated in Experiment 1 were recruited from the Brown University community using the previously outlined criteria. Half of the subjects were assigned to the syntactic encoding condition, and the other half were assigned to the talker identification encoding condition. An additional five listeners participated but were excluded from the analyses because they did not meet the criterion for hit rate in the recognition phase.

Stimuli and design The stimuli and design used in Experiment 1 were also those used in Experiment 2.

Procedure The procedure outlined for Experiment 1 was the same one used for Experiment 2, with one exception: In this experiment, attention during encoding was directed to either syntactic information or talker identity. Listeners in the syntactic encoding condition were asked to listen to each word presented during the encoding phase and to decide, on each trial, whether the word was only a noun (e.g., *cat*) or was or could be another part of speech (e.g., *sad*, *bat*). Listeners made their decision by pressing one of two buttons labeled "noun only" and "not noun only." As in Experiment 1, the listeners in the talker identification encoding condition were asked to indicate talker gender on each trial by pressing one of two buttons labeled "male" and "female." A brief pause (2–3 min) was interposed between the encoding and recognition phases.

Results

Hit rate Hit rates were analyzed as outlined in Experiment 1. The left panel of Fig. 2 shows the mean hit rates for same-talker and different-talker trials during the recognition phase for listeners in the syntactic and talker identification encoding conditions. The results of a two-way ANOVA revealed a main effect of recognition condition [$F(1, 22) = 23.42$, $p < .001$, $\eta^2 = .515$], no main effect of trial type [$F(1, 22) = 0.16$, $p = .696$, $\eta^2 = .008$], and no interaction between recognition condition and trial type [$F(1, 22) = 0.02$, $p = .896$, $\eta^2 = .000$]. The main effect of encoding condition reflected a higher hit rate in the syntactic encoding condition (mean = .93) than in the talker identification encoding condition (mean = .78). These results indicate that listeners who attended to syntactic information during encoding showed better recognition memory for words than did listeners who attended to talker gender during encoding. However, neither group of listeners showed a talker-specificity effect; hit rates were equivalent for same-talker and different-talker trials in both groups of listeners.

Reaction time RTs were analyzed as outlined in Experiment 1. Fifteen data points were outliers (3.7 % of the total data) and were removed from subsequent analyses. The right panel of Fig. 2 shows the mean RTs during recognition for the two encoding conditions for same-talker and different-talker trials. Mean RTs were submitted to an ANOVA with the between-subjects factor Encoding Condition (syntactic decision, talker identification) and the within-subjects factor Trial Type (same talker, different talker). The results showed a main effect of encoding condition [$F(1, 22) = 10.25$, $p = .004$, $\eta^2 = .318$],

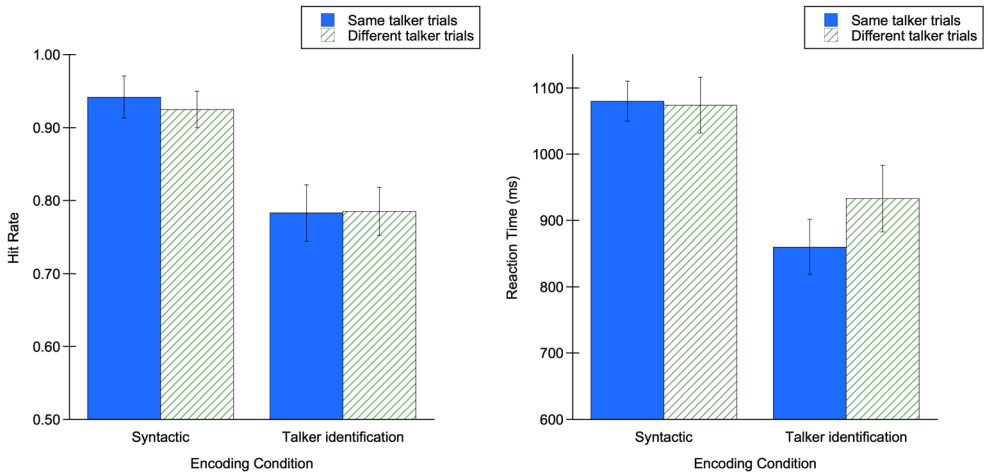

Fig. 2 Mean hit rates (left panel) and reaction times (in milliseconds, right panel) for hits during the recognition phase of Experiment 2, for each encoding condition and for same-talker and different-talker trials. Error bars indicate standard errors of the means

with the mean RT in the syntactic encoding condition being substantially longer than the mean RT in the talker identification encoding condition (1077 vs. 897 ms, respectively). We observed no main effect of trial type [$F(1, 22) = 4.12$, $p = .06$, $\eta^2 = .129$]. However, there was a significant interaction between encoding condition and trial type [$F(1, 22) = 5.73$, $p = .026$, $\eta^2 = .180$]. Planned comparisons revealed that the interaction was due to faster RTs to same-talker than to different-talker trials in the talker identification encoding condition [860 vs. 933 ms, respectively; $t(11) = -2.95$, $p = .013$, $d = 0.047$], but the RTs to same-talker and different-talker trials were equivalent in the syntactic encoding condition [1080 vs. 1074 ms, respectively; $t(11) = .028$, $p = .787$, $d = -0.455$].

Discussion

When attention during encoding was specifically directed toward talker identity, a talker-specificity effect emerged during recognition, such that listeners were faster to respond to same-talker than to different-talker trials. No talker-specificity effect was observed at recognition when attention during encoding was directed toward syntactic information. These findings are consistent with the results from Experiment 1 and, moreover, suggest that the specificity effect observed in Experiment 1 was due to attention during encoding and was not simply the consequence of making an overt decision during encoding. We note, however, that unlike in Experiment 1, in which specificity effects were observed for both the hit rate and RT analyses, in Experiment 2 an effect was only observed for the RT data. This finding suggests that RT may be a more sensitive measure of specificity effects than is hit rate, at least in this paradigm, and that the specificity effect observed for hit rates in Experiment 1 should be interpreted cautiously, given that it did not replicate in Experiment 2.

The emergence of a talker-specificity effect in the RT data cannot be attributed to an increase in processing time, because RTs at recognition were far longer for the group of listeners who made syntactic decisions during encoding than for the group who made talker decisions. This pattern of results is not consistent with the time-course hypothesis, which predicts that the specificity effect should have emerged for the syntactic group, who had relatively slower processing times during retrieval. As in Experiment 1, we examined processing times during the encoding phase. The mean RT for syntactic decisions was significantly longer than that for talker decisions [2258 vs. 997 ms, respectively; $t(22) = 11.03$, $p < .0001$, $d = 4.505$], as we would expect on the basis of earlier work showing processing delays (and also higher hit rates) that were associated with increased depth of processing (Craik & Tulving, 1975). Thus, even though listeners in the syntactic encoding condition had longer processing times during both encoding and recognition than did those in the talker identification condition, they did not show talker-specificity effects during recognition.

Experiment 3

The results from Experiments 1 and 2 have demonstrated that manipulating attention during encoding can influence the emergence of specificity effects during subsequent recognition. Moreover, these attention-driven specificity effects occurred in the absence of a concomitant increase in processing time. The goal of Experiment 3 was to provide an additional test of the time-course hypothesis by specifically manipulating processing time while meeting three constraints for the "fast" and "slow" conditions: (1) the same stimuli must be used, (2) attention must be held constant, and (3) task difficulty must not differ between the two conditions. To this end, two groups of listeners participated in a recognition memory experiment consisting of an encoding phase and a recognition phase. The encoding phases were identical for both groups; listeners were asked simply to listen to a series of words. Accordingly, attention

for both groups of listeners was directed to general lexical characteristics, as had been the case for one group of listeners in Experiment 1. The recognition phases differed across the two groups, in that half of the listeners performed the recognition task in quiet and the other half performed the task in the context of background noise. We expected that response latencies would be substantially longer when processing speech in noise than in quiet, even at the favorable signal-to-noise ratio employed in this experiment. As we discuss in detail in the Summary and Conclusions section, manipulating processing time independently of task difficulty is no mean feat. However, as we describe below, the manipulation used here was selected because it allowed for equivalent hit rates (a metric of difficulty) between the "fast" and "slow" conditions.

If specificity effects in spoken word recognition solely reflect the point in time at which a particular representation is retrieved, as predicted by the time-course hypothesis, then specificity effects should emerge for listeners in the noise condition but not in the quiet condition, in line with the slower processing times expected in the noise condition. If attention during encoding is the central determinant of specificity effects during recognition, as was suggested by the results of Experiments 1 and 2, then we would predict that talker-specificity effects should fail to emerge for both listeners groups, despite differences in processing time, given that both groups' attention during the encoding phase was directed toward general lexical characteristics and not to talker characteristics.

Method

Subjects Twenty-four subjects who had not participated in Experiments 1 or 2 were recruited from the Brown University community using the previously outlined criteria. Half of the subjects were assigned to the recognition-in-quiet condition, and the other half were assigned to the recognition-in-noise condition. An additional three listeners participated but were excluded from the analyses because they did not meet the criterion for recognition hit rate.

Stimuli and design The same stimuli and design used in Experiments 1 and 2 were also used in Experiment 3.

Procedure The procedure outlined for Experiment 1 was the same one used for Experiment 3, with two exceptions. First, encoding for both groups of listeners followed the format of the lexical condition described in Experiment 1. That is, all listeners were directed to listen to each word presented during encoding and to press a button to advance to the next trial. This condition was thus identical to that used in Mattys and Liss (2008). Second, half of the listeners completed the recognition phase in quiet, and the other half completed the recognition phase in background noise. The noise was a slightly modified version of the multitalker babble developed for the Speech Perception in Noise test (Kalikow, Stevens, & Elliot, 1977). As in Experiments 1 and 2, auditory stimuli were presented at 59 dB SPL. The noise was presented at 63 dB SPL, which yielded a signal-to-noise ratio of –4 dB SPL.

Results

Hit rate Hit rates were analyzed as outlined for Experiments 1 and 2. The left panel of Fig. 3 shows the mean hit rates for same-talker and different-talker trials for listeners in the quiet and noise recognition conditions. The results of a two-way ANOVA showed no main effect of recognition condition [$F(1, 22) = 3.16$, $p = .089$, $\eta^2 = .126$], no main effect of trial type [$F(1, 22) = 0.15$, $p = .701$, $\eta^2 = .071$], and no interaction between recognition condition and trial type [$F(1, 22) = 0.02$, $p = .898$, $\eta^2 = .000$]. These results indicate that hit rates were statistically equivalent across the two recognition conditions and that neither group showed a specificity effect.

Reaction time RTs were analyzed as outlined for Experiments 1 and 2. Twenty-five of the data points were outliers (6.5 % of the total data) and were removed from subsequent analyses. The right panel of Fig. 3 shows the mean RTs for the two recognition conditions for same-talker and different-talker trials. Mean RTs were submitted to an ANOVA with the between-subjects factor Recognition Condition (quiet, noise) and the within-subjects factor Trial Type (same talker, different talker). The results showed a main effect of condition [$F(1, 22) = 4.37$, $p = .048$, $\eta^2 = .166$], with the RT in the noise condition being 140 ms longer than the RT in the quiet condition (1041 vs. 901 ms, respectively). We found no main effect of trial type [$F(1, 22) = 0.40$, $p = .536$, $\eta^2 = .017$], indicating that the mean RT for same-talker trials was equivalent to that for different-talker trials. Moreover, there was no interaction between condition and trial type [$F(1, 22) = 0.26$, $p = .619$, $\eta^2 = .011$].[2]

[2] As in Experiments 1 and 2, we analyzed the RTs during encoding for both groups of listeners in Experiment 3 (measured from the onset of the auditory stimulus to the onset of the buttonpress to advance to the next word). The difference in encoding processing times between the two groups (recognition in quiet vs. in noise) was not statistically reliable, as expected, given that the encoding conditions for both groups were identical [1056 vs. 1469 ms, respectively; $t(22) = -1.14$, $p = .266$, $d = -0.463$]. Nonetheless, there was a large numerical difference in mean RTs. Inspection of the data revealed one outlier subject who did not respond to any trial during the encoding phase. Thus, the next trial advanced only after the response timeout of 5000 ms had been reached, and this extremely long RT was recorded for each trial. To ensure that the lack of a significant difference between the two groups was not due to extreme variability as a result of this participant, we compared the mean RTs between the two groups after removing this participant. We again observed no significant difference in encoding processing times for the listeners who performed the recognition task in quiet versus noise [1056 vs. 1148 ms, respectively; $t(21) = -0.52$, $p = .605$, $d = -0.219$].

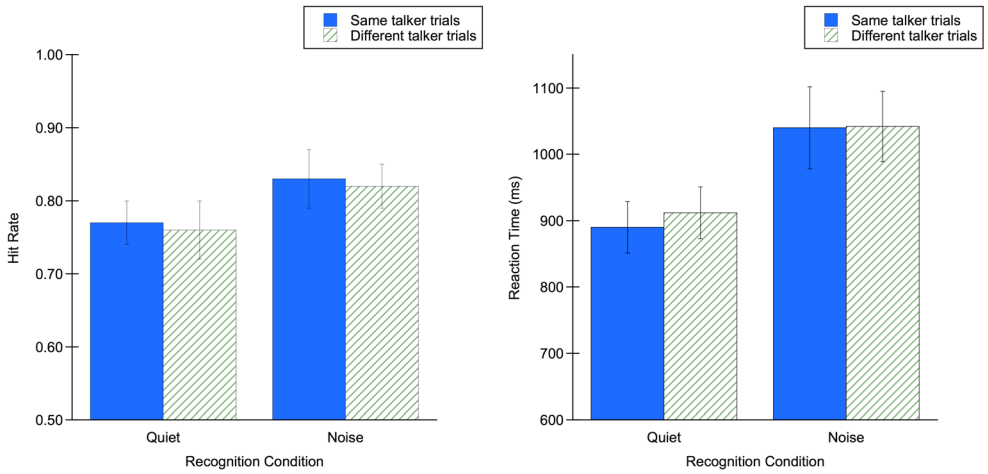

Fig. 3 Mean hit rates (left panel) and reaction times (in milliseconds, right panel) for hits during the recognition phase of Experiment 3, for each recognition condition and for same-talker and different-talker trials. Error bars indicate standard errors of the means

Discussion

The mean processing time for listeners who performed the recognition task in background noise was 140 ms slower than that for listeners who performed the same task in quiet. The magnitude of this RT difference was greater than that shown in previous studies that had examined specificity effects as a function of processing time (McLennan & Luce, 2005). Despite the increased processing time in background noise, no evidence for a specificity effect was found in the "slow" condition. The lack of a specificity effect despite slower RTs in the noise condition suggests that listeners relied on abstract information during recognition, and not specific episodic traces, as predicted by the time-course hypothesis. These data suggest that when task difficulty is held constant, processing time fails to predict the emergence of specificity effects.

Summary and conclusions

A host of findings have indicated that listeners have access to both abstract and episodic information within the language architecture. As a case in point, listeners readily comprehend the speech of unfamiliar talkers. However, given experience with a particular talker, talker familiarity effects are robust and are observed at both prelexical (Theodore & Miller, 2010) and lexical (McLennan & Luce, 2005) levels of processing. Thus, a complete model of spoken language comprehension must specify the factors that influence when listeners will recruit one source of information over the other. One prominent theory is formalized in the time-course hypothesis (McLennan & Luce, 2005). The primary assumption behind this hypothesis is that abstract information, such as a summary representation or allophonic variation, is far more frequent in its representation than is episodic information, such as the acoustic trace associated with a particular talker's production. A secondary assumption is that more-frequent representations require less time to reach threshold for activation than do less-frequent representations. Accordingly, the time-course hypothesis predicts that specificity effects associated with episodic information will emerge only late in the processing stream, with abstract representations prevailing when recognition occurs relatively earlier.

As we reviewed in the introduction, the evidence to date in support of the time-course hypothesis has not completely distinguished between processing time and other factors that may influence the use of abstract versus episodic information, such as task difficulty, attention, and the very stimuli presented to listeners. As a consequence, what has been attributed to differences in the time courses of lexical retrieval may actually have been due to encoding factors such as attention or depth of processing. In the present work, we aimed to examine the predictions of the time-course hypothesis in cases in which encoding factors were manipulated orthogonally to retrieval factors. Our results failed to support the predictions of the time-course hypothesis. Experiment 1 showed that when attention was directed toward talker identity, talker-specificity effects emerged even when recognition occurred early in the processing stream. The results of Experiment 2 provided further support for the role of attention in mediating talker-specificity effects, such that the effect is not solely the consequence of depth of processing during encoding; rather, attention must be specifically directed toward talker identity. The results of Experiment 3, which held the stimulus set constant across "fast" and "slow" conditions, demonstrated that simply delaying lexical retrieval by means of the addition of background noise is not sufficient to promote reliance on episodic information.

In moving forward, the results of the present experiments point to two critical considerations for the time-course hypothesis of specificity effects in spoken word recognition. First, one challenge for the hypothesis will be to operationally

define early versus late processing. On the basis of previous research, it is not clear what absolute difference in processing time would be required to allow for access to episodic information. In the lexical decision paradigm used by McLennan and Luce (2005), the presence of a specificity effect depended on a processing time difference of as little as 35 ms. In contrast, the difference between "fast" (normal-speech condition) and "slow" (dysarthric-speech condition) processing in the recognition memory paradigm used by Mattys and Liss (2008) was around 200 ms. A second challenge for the time-course hypothesis will be to provide an architecture that would allow for encoding factors, such as attention, to be examined independently of retrieval factors, such as processing time. As it is currently implemented, this hypothesis posits that a retrieval mechanism is the primary determinant between the use of abstract versus episodic information. The results from the present study suggest that attention during encoding not only predicts when each source of information will be used, but that it does so even when pitted against processing time during retrieval. Models of spoken word recognition therefore need to include a role for attention in modulating specificity effects. Here we considered attention specifically during encoding, and future work should also consider the role of attention during retrieval. Attention, as it is broadly characterized in cognitive psychology, modulates the resources devoted to information processing, including encoding and retrieving the sensory properties of the stimulus. As a consequence, attention may serve to increase the salience of the attended properties of a representation, resulting in increased activation of episodic traces without a requisite increase in processing time.

Author note This research was supported by Grant No. R01 DC000314 from the National Institutes of Health (NIH), National Institute on Deafness and Other Communication Disorders (NIDCD). The content is the responsibility of the authors and does not necessarily represent the official views of the NIH or the NIDCD. Portions of these data were presented at the 162nd meeting of the Acoustical Society of America. We extend gratitude to Conor T. McLennan, Sven L. Mattys, and two anonymous reviewers for their helpful comments on a previous version of the manuscript.

Appendix

bad	dig	hip	nap	sad
bat	fan	jam	net	sag
book	fed	jet	nut	sin
bug	gas	leg	pen	sip
bus	goat	map	pet	tub
cab	hat	mat	pig	van
cat	hem	mop	pill	wed
cup	hen	nail	ran	win

References

Boersma, P., & Weenink, D. (2011). Praat: Doing phonetics by computer [Computer program]. Retrieved from www.praat.org

Church, B. A., & Schacter, D. L. (1994). Perceptual specificity of auditory priming: Implicit memory for voice intonation and fundamental frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 521–533. doi:10.1037/0278-7393.20.3.521

Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *Journal of the Acoustical Society of America, 116,* 3647–3658.

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104,* 268–294. doi:10.1037/0096-3445.104.3.268

Delattre, P. C., Liberman, A. M., & Cooper, F. S. (1955). Acoustic loci and transitional cues for consonants. *Journal of the Acoustical Society of America, 27,* 769–773.

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. doi:10.1037/0278-7393.22.5.1166

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105,* 251–279. doi:10.1037/0033-295X.105.2.251

Grossberg, S. (1986). The adaptive self-organization of serial order in behavior: Speech, language, and motor control. In E. C. Schwab & H. C. Nusbaum (Eds.), *Pattern recognition by humans and machines (Vol. 1): Speech perception* (pp. 187–294). New York, NY: Academic Press.

Kalikow, D. N., Stevens, K. N., & Elliot, L. L. (1977). Development of a test 80 of speech intelligibility in noise using sentence materials with controlled word predictability. *Journal of the Acoustical Society of America, 61,* 1337–1351.

Klatt, D. H. (1986). The problem of variability in speech recognition and in models of speech perception. In J. S. Perkell & D. H. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 300–319). Hillsdale, NJ: Erlbaum.

Krestar, M. L., & McLennan, C. T. (2013). Examining the effects of variation in emotional tone of voice on spoken word recognition. *Quarterly Journal of Experimental Psychology, 66,* 1793–1802. doi:10.1080/17470218.2013.766897

Ladefoged, P., & Broadbent, D. E. (1957). Information conveyed by vowels. *Journal of the Acoustical Society of America, 29,* 98–104.

Magnuson, J. S., & Nusbaum, H. C. (2007). Acoustic differences, listener expectations, and the perceptual accommodation of talker variability. *Journal of Experimental Psychology: Human Perception and Performance, 33,* 391–409. doi:10.1037/0096-1523.33.2.391

Mattys, S. L., & Liss, J. M. (2008). On building models of spoken-word recognition: When there is as much to learn from natural "oddities" as artificial normality. *Perception & Psychophysics, 70,* 1235–1242. doi:10.3758/PP.70.7.1235

McLennan, C. T., & González, J. (2012). Examining talker effects in the perception of native- and foreign-accented speech. *Attention, Perception, & Psychophysics, 74,* 824–830. doi:10.3758/s13414-012-0315-y

McLennan, C. T., & Luce, P. A. (2005). Examining the time course of indexical specificity effects in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 306–321. doi:10.1037/0278-7393.31.2.306

McLennan, C. T., Luce, P. A., & Charles-Luce, J. (2003). Representation of lexical form. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 539–553. doi:10.1037/0096-1523.29.4.539

Miller, J. L. (1981). Effects of speaking rate on segmental distinctions. In P. D. Eimas & J. L. Miller (Eds.), *Perspectives on the study of speech* (pp. 39–74). Hillsdale, NJ: Erlbaum.

Mullennix, J. W., Pisoni, D. B., & Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America, 85,* 365–378.

Munro, M. J. (1998). The effects of noise on the intelligibility of foreign-accented speech. *Studies in Second Language Acquisition, 20,* 139–153.

Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehension, and intelligibility in the speech of second language learners. *Language Learning, 45,* 73–97.

Nygaard, L. C., Burt, S. A., & Queen, J. S. (2000). Surface form typicality and asymmetric transfer in episodic memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 1228–1244. doi:10.1037/0278-7393.26.6.1228

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science, 5,* 42–46.

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. doi:10.1037/0278-7393.19.2.309

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America, 24,* 175–184.

Schacter, D. L., & Church, B. A. (1992). Auditory priming: Implicit and explicit memory for words and voices. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 915–930. doi: 10.1037/0278-7393.18.5.915

Theodore, R. M., & Miller, J. L. (2010). Characteristics of listener sensitivity to talker-specific phonetic detail. *Journal of the Acoustical Society of America, 128,* 2090–2099.

Theodore, R. M., Miller, J. L., & DeSteno, D. (2009). Individual talker differences in voice-onset-time: Contextual influences. *Journal of the Acoustical Society of America, 125,* 3974–3982.