Published in final edited form as:
Atten Percept Psychophys. 2015 July ; 77(5): 1674–1684. doi:10.3758/s13414-015-0854-0.

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

Rachel M. Theodore[1], **Sheila E. Blumstein**[2], and **Sahil Luthra**[2]

[1]Department of Speech, Language, and Hearing Sciences, University of Connecticut

[2]Department of Cognitive, Linguistic, and Psychological Sciences, Brown University

Abstract

Findings in the domain of spoken word recognition indicate 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 required to access lower-frequency episodic instantiations. The time-course hypothesis of specificity effects thus identifies a strong role for retrieval mechanisms mediating the use of abstract versus episodic information. Here we conducted three recognition memory experiments to examine whether findings previously attributed to retrieval mechanisms might reflect attention during encoding. Results from Experiment 1 showed that talker-specificity effects emerged when subjects attended to individual speakers during encoding, but not when they attended to lexical characteristics during encoding, even though processing time at retrieval was equivalent. 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 processing time at retrieval was significantly longer in the latter condition. Results from Experiment 3 showed no talker-specificity effects when attending 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.

Introduction

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 representation. 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 et al., 1955), and even idiosyncratic differences in pronunciation across individual talkers (e.g., Klatt, 1986; Theodore et al.,

Address correspondence to Rachel M. Theodore, Department of Speech, Language, and Hearing Sciences, University of Connecticut, Unit 1085, Storrs, CT 06269-1085, rachel.theodore@uconn.edu.

2009). Given this variability, the challenge for the listener is to recognize physically distinct objects as 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 et al., 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 suggest that listeners do retain surface characteristics for individual words (Goldinger, 1998; Palmeri et al., 1993), which supports episodic-based models that posit 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, but 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 compared to unfamiliar voices (Nygaard et al., 1994), but processing time is faster for familiar compared to 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 suggest that such talker-specificity effects, while 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 et al., 2003). McLennan and colleagues explain this difference in terms of the relative frequency of both types of variability. They posit that allophonic variability, such as a flap produced for medial /t/, is more frequently encountered than any particular talker's phonetic signature. They model 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 comes 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. Response time to the impaired speech was longer than for 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. 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 speech compared to 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, then the lower frequency, episodic information will prevail and performance will show specificity effects.

The initial examination of the time-course hypothesis (McLennan & Luce, 2005) used task difficulty to manipulate processing time, with an easy task used to generate “fast” processing times and a difficult task used to generate relative slower processing times. For example, listeners completed a lexical decision task where nonwords were either very similar to real words and thus were difficult to identify as nonwords or they were maximally distinct from real words and thus were easily identified as nonwords. Other tasks used to manipulate processing time included immediate versus delayed shadowing, where 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 processing time per se.

Rather than explicitly manipulating processing time through task difficulty, Mattys and Liss (2008) manipulated processing time by varying the nature of the stimuli presented for the fast versus slow conditions. Stimuli in the fast condition consisted of typical speech and stimuli in the slow condition consisted of dysarthric speech, which may have been an implicit manipulation of task difficulty. Indeed, those who heard dysarthric speech had much lower hit rate compared to those who heard typical speech, suggesting that processing dysarthric speech was much more difficult than processing the typical speech. The specificity effect for the dysarthric speech was observed even when analyzing only those items that were correctly identified in intelligibility pre-tests, which indicates that it 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 comes from McLennan and Gonzalez (2012) who examined 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 is 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 listeners who heard native speech. Given the literature demonstrating increased difficulty in processing foreign-accented speech compared 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 source of information used to guide a particular response. Hence, the current work seeks to evaluate the time-course hypothesis in the case where processing time is decoupled from task difficulty.

To date, the literature on the time-course hypothesis has focused on processing time at retrieval. However, there is a large body of evidence indicating that 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 previous findings attributed to differences in processing time during retrieval have actually been the consequence of differences in encoding factors. The current work tests this hypothesis. We use the recognition memory paradigm of Mattys and Liss (2008) to examine the role of attention during encoding on 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 (Experiment 1) or syntactic (Experiment 2) aspects of the signal. Following encoding, all participants completed a recognition memory test where they were asked to indicate on each trial whether they had heard that word during encoding. In Experiment 3, attention during encoding was directed towards 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 hit rate and response time at recognition were influenced by whether voice was held constant for a given word between the encoding and recognition phases. If, as predicted by the time-course hypothesis, specificity effects associated with the use of episodic information are determined by processing time during retrieval, then we should only observe a specificity effect for listeners who complete the recognition task in background noise and thus have the slowest processing times. If, however, specificity effects reflect the role of attention during encoding, then we will observe specificity effects only when listeners attend to talker identity, and they will emerge irrespective of processing time such and be observed even when processing is 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 phase was identical for both groups of listeners and is a replication of 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 directing 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 where attention during encoding was manipulated orthogonally to processing time during retrieval and, critically, to do so in a “fast” condition. Because the recognition phase is identical for both encoding groups, we predict that reaction times during recognition will 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 are elicited early in the processing stream in a “fast” condition. If, however, attention during encoding influences subsequent recognition memory, then we predict that talker-specificity effects will be observed for listeners who attended to talker characteristics during encoding, despite equivalent (and fast) reaction times compared to listeners who attended to general lexical characteristics.

Method

Participants—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 analyses because they did not meet the criterion for recognition hit rate, as described below.

Stimuli—Stimuli included 40 monosyllabic words with consonant-vowel-consonant syllable structure and are listed in the Appendix. Words were selected to be familiar, 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, mean fundamental frequency for the female talker was 185 Hz (SD = 28) and mean fundamental frequency for the male talker was 114 Hz (SD = 19). Mean word duration for the female talker was 474 ms (SD = 66) and 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 were previously presented during encoding (“old” words) and 10 words were 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 an encoding-recognition pair, with voice 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 an equal number 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, order of presentation of items for the encoding and recognition phase 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. 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 subjects completed an encoding phase followed by a recognition phase. Listeners in the lexical encoding condition were instructed to listen carefully to each word and press a button to advance to the next word. Listeners in the talker identification encoding condition were instructed to listen carefully to each word and indicate the gender of the talker by pressing the appropriately labeled button on the response box. The recognition phase was identical for listeners in both encoding conditions; all were directed to indicate on each trial whether or not the word was presented during encoding by pressing a button labeled "yes" or "no." 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 encoding and recognition phases, the pause between trials was 2000 ms, timed from the button response. There was a very short break (approximately 2–3 minutes) between the two phases.

Results

Hit rate—We analyzed performance during the recognition phase for both groups of listeners as follows. Mean hit rate was 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 a hit rate greater than .60 for both same talker and different talker trials. Two subjects were replaced because they failed to meet this criterion.

Figure 1 (left panel) shows mean hit rate across listeners for same talker and different talker trials separately for each encoding condition. Mean hit rate was submitted to an ANOVA with the between-subjects factor of encoding condition (lexical, talker identification) and the within-subjects factor of 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 indicating that directing listeners to attend to the word or to the talker did not influence overall recognition memory. However, there was 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 ($\alpha = 0.025$). The results showed that the interaction was due to the hit rate for same talker trials being significantly higher than different talker trials for the talker identification encoding condition [0.88 vs. 0.78, respectively, t(11) = 2.71, p = .020, *d* = 0.989], but not for the lexical encoding condition [0.79 vs. 0.81, respectively, t(11) = 0.56, p = .586, *d* = −0.149].[1]

Reaction time—Reaction time (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, different talker trials were considered outliers and removed from subsequent analysis. Sixteen data points (4.1% of total data) were removed for this reason. The right panel of Figure 1 shows the mean RT 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 RTs were equivalent across the two listener groups [F(1,22) = 0.43, p = .518; η^2 = .019]. There was a marginal main effect of trial type [F(1,22) = 3.76, p = .066; η^2 = .124] and a significant interaction between trial type and encoding condition [F(1,22) = 4.615, p = .043; η^2 = .152]. Planned comparisons showed that the interaction was due to faster RTs to same talker compared to different talker trials in the talker identification condition [906 ms vs. 977 ms, respectively, t(11) = 2.74, p = .019, *d* = −.0543], but not in the lexical condition [985 ms vs. 981 ms, respectively, t(11) = 0.16, p = .877, *d* = 0.021].

Discussion

When attention was explicitly directed towards talker characteristics during the encoding phase, listeners demonstrated a processing advantage for recognition of words that were presented in the same voice between encoding and recognition compared to when voice differed across the two phases. This specificity effect indicates that listeners relied on specific episodic representations to facilitate lexical recognition. In contrast, listeners who attended to more general lexical characteristics during encoding did not show a specificity effect during recognition. Processing time at recognition for both groups of listeners was 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 time did not differ between the two listener groups, as measured by RT 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 it was not attention to talker identity *per se* that gave rise to the specificity effects, but rather that it was the consequence of requiring participants to make a decision during encoding that led to specificity effects at recognition. Recall that listeners in the talker identification group were required to make a talker gender decision on

[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 if performance in the experiment differed as a function of the two talkers' voices. For these analyses, mean hit rate and mean reaction time were submitted to repeated-measures ANOVA with the factors of 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 (p > .10 in all cases). Accordingly, we collapsed across talker in order to perform the analyses presented in the main text.

every trial during encoding; in contrast, listeners in the lexical encoding condition were directed to listen to each word and press a button to advance to the next trial. This could have potentially led to a situation where 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 listeners in the lexical encoding condition were not actually attending to lexical characteristics as we had intended, but were simply pressing a button to move the next trial. To address this possibility, we analyzed reaction times during encoding (measured from the onset of the auditory stimulus to the onset of the button press) and found that processing time was significantly longer in the lexical compared to the talker identification encoding condition [1471 ms 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. Experiment 2 directly examines this possibility.

Experiment 2

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 towards talker identity but was not observed when attention was directed towards general lexical characteristics. In order to ensure that this pattern of results is 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), Experiment 2 examines the role of attention in a case where listeners are always required to make a decision during encoding.

Two groups of listeners participated in encoding and recognition phases similar in design to 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 an identical recognition memory task as described for Experiment 1. If, as suggested by the results of Experiment 1, attention during encoding influences the emergence of specificity effects at recognition, then we predict that a talker-specificity effect will only be observed for those who attended to talker identity. If however, results from Experiment 1 reflect the consequence of making a judgment during encoding irrespective of attention demands, then we predict that a talker-specificity effect will emerge for both groups of listeners.

Method

Participants—Twenty-four subjects who did not participate 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 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 used in Experiment 2.

Procedure—The procedure outlined for Experiment 1 was the same 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 decide, on each trial, whether the word was only a noun (e.g., *cat*) or whether the word 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, 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.” There was a brief pause (2–3 minutes) between the encoding and recognition phases.

Results

Hit rate—Hit rate was analyzed as outlined in Experiment 1. The left panel of Figure 2 shows mean hit rate for same talker and different talker trials during the recognition phase for listeners in the syntactic and talker identification encoding conditions. Results of a 2-way ANOVA revealed a main effect of recognition condition [F(1,22) = 23.42, p < .001; η^2 = .515], no main effect of trial type [F(1,22) = 0.16, p = .696; η^2 = .008], and no interaction between recognition condition and trial type [F(1,22) = 0.02, p = .896; η^2 = .000]. The main effect of encoding condition reflected higher hit rate in the syntactic encoding condition (mean = 0.93) compared to the talker identification encoding condition (mean = 0.78). These results indicate that listeners who attended to syntactic information during encoding showed better recognition memory for words compared to listeners who attended to talker gender during encoding. However, neither group of listeners showed a talker-specificity effect; hit rate was equivalent for same talker and different talker trials in both groups of listeners.

Reaction time—Reaction time was analyzed as outlined in Experiment 1. Fifteen data points were outliers (3.7% of the total data) and removed from subsequent analyses. The right panel of Figure 2 shows mean RT during recognition for the two encoding conditions for same talker and different talker trials. Mean RT was submitted to an ANOVA with the between-subjects factor of encoding condition (syntactic decision, talker identification) and the within-subjects factor of trial type (same talker, different talker). Results showed a main effect of encoding condition [F(1,22) = 10.25, p = .004; η^2 = .318], with mean RT in the syntactic encoding condition substantially longer than mean RT in the talker identification encoding condition (1041 ms versus 901 ms, respectively). There was no main effect of trial type [F(1,22) = 4.12, p = .06; η^2 = .129]. However, there was a significant interaction between encoding condition and trial type [F(1,22) = 5.73, p = .026; η^2 = .180]. Planned comparisons revealed that the interaction was due to faster RTs to same talker compared to different talker trials in the talker identification encoding condition [856 ms vs. 953 ms, respectively; t(11) = −2.95, p = .013; *d* = .047], but RTs to same and different talker trials was equivalent in the syntactic encoding condition [1080 ms vs. 1074 ms, respectively, t(11) = .028, p = .787; *d* = −.455].

Discussion

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

The emergence of a talker-specificity effect in the reaction time data cannot be attributed to an increase in processing time because reaction times at recognition were far longer for the group of listeners who made syntactic decisions during encoding compared to the group of listeners 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 time during retrieval. As in Experiment 1, we examined processing time during the encoding phase. Mean reaction time for syntactic decisions was significantly longer compared to talker decisions [2258 ms vs. 997 ms, respectively, $t(22) = 11.03$, $p < .0001$, $d = 4.505$], as expected based on earlier work showing processing delays (and also higher hit rates) 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 compared to those in the talker identification condition, they did not show talker-specificity effects during recognition.

Experiment 3

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 phase was identical for both groups; listeners were asked to simply listen to a series of words. Accordingly, attention for both groups of listeners was directed to general lexical characteristics, as was the case for one group of listeners in Experiment 1. The recognition phase 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 latency would be substantially longer when processing speech in noise compared to quiet, even at the favorable signal-to-noise ratio employed in this experiment. As discussed in detail in Summary and Conclusions, manipulating processing time independently of task difficulty is no mean feat. However, as described below, the manipulation used here was selected because it allowed for equivalent hit rate (a metric of difficulty) between the "fast" and "slow" conditions.

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

Method

Participants—Twenty-four subjects who did not participate 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 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 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 press a button to advance to the next trial. This condition is 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 multi-talker babble developed for the Speech Perception in Noise test (Kalikow et al., 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 rate was analyzed as outlined for Experiments 1 and 2. The left panel of Figure 3 shows mean hit rate for same talker and different talker trials for listeners in the quiet and noise recognition conditions. Results of a 2-way ANOVA showed no main effect of recognition condition [F(1,22) = 3.16, p = .089; η^2 = .126], no main effect of trial type [F(1,22) = 0.15, p = .701; η^2 = .071], and no interaction between recognition condition and trial type [F(1,22) = 0.02, p = .898; η^2 = .000]. These results indicate that hit rate was statistically equivalent across the two recognition conditions and that neither group showed a specificity effect.

Reaction time—Reaction time was analyzed as outlined for Experiments 1 and 2. Twenty-five data points were outliers (6.5% of the total data) and removed from subsequent analyses. The right panel of Figure 3 shows mean RT for the two recognition conditions for same talker and different talker trials. Mean RT was submitted to an ANOVA with the between-subjects factor of recognition condition (quiet, noise) and the within-subjects factor or trial type (same talker, different talker). Results showed a main effect of condition [F(1,22) = 4.37, p = .048; η^2 = .166], with RT in the noise condition 140 ms longer than RT in the quiet condition (1041 ms versus 901 ms, respectively). There was no main effect of trial type [F(1,22) = 0.40, p = .536; η^2 = .017], indicating that 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; η^2 = .011].[2]

Discussion

Mean processing time for listeners who performed the recognition task in background noise was 141 ms slower compared to listeners who performed the same task in quiet. The magnitude of this RT difference was greater than that shown in previous studies examining specificity effects as a function of processing time (McClennan and 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

There are a host of findings indicating 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 prelexical (Theodore & Miller, 2010) and lexical levels of processing (McLennan & Luce, 2005). 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

[2] As in Experiments 1 and 2, we analyzed reaction time during encoding for both groups of listeners in Experiment 3 (measured from the onset of the auditory stimulus to the onset of button press to advance to the next word). The difference in encoding processing time between the two groups (recognition in quiet versus recognition in noise) was not statistically reliable, as expected, given that the encoding condition for both groups was identical [1056 ms vs. 1469 ms, respectively, t(22) = −1.14, p = .266, *d* = −0.463]. Nonetheless, there was a large numerical difference in mean reaction time. Inspection of the data revealed one outlier participant who did not respond to any trial during the encoding phase. Thus, the next trial advanced only after the response time-out of 5000 ms was reached and this extremely long reaction time 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 mean reaction time between the two groups removing this participant. We again observed no significant difference for encoding processing time for the listeners who performed the recognition task in quiet versus noise [1056 ms vs. 1148 ms, respectively, t(21) = −0.52, p = .605, *d* = −0.219].

assumption behind this hypothesis is that abstract information, such as a summary representation or allophonic variation, is far more frequent in its representation than episodic information, such as an 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 less frequent representations. Accordingly, the time-course hypothesis predicts that specificity effects associated with episodic information will only emerge late in the processing stream, with abstract representations prevailing when recognition occurs relatively earlier.

As reviewed in the Introduction, evidence to date in support of the time-course hypothesis does not completely distinguish 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 time-course of lexical retrieval may actually have been due to encoding factors such as attention or depth of processing. The current work aimed to examine predictions of the time-course hypothesis in cases where encoding factors were manipulated orthogonally to retrieval factors. Results failed to support the predictions of the time course hypothesis. Experiment 1 showed that when attention was directed towards talker identity, talker-specificity effects emerged even when recognition occurred early in the processing stream. Results of Experiment 2 provided further support for the role of attention in mediating talker-specificity effects such that it is not solely the consequence of depth of processing during encoding; rather, attention must be specifically directed towards talker identity. Results of Experiment 3, which held the stimulus set constant across “fast” and “slow” conditions, demonstrated that simply delaying lexical retrieval through the addition of background noise is not sufficient to promote reliance on episodic information.

In moving forward, the results of the current experiments point to two critical considerations for the time-course hypothesis of specificity effects in spoken word recognition. First, one challenge for the time-course hypothesis is to operationally define early versus late processing. Based on 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 (2008), 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 was around 200 ms. A second challenge for the time-course hypothesis is 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. Results from the current 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 broadly characterized in cognitive psychology, modulates resources devoted to information processing, including encoding and retrieving sensory properties of the stimulus. As a consequence, attention may serve to increase salience of the attended properties of a representation, resulting in increased activation of episodic traces without a requisite increase in processing time.

Acknowledgments

This research was supported by Grant 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 official views of the NIH or the NIDCD. Portions of these data were presented at the 162nd meeting of the Acoustical Society of America. Gratitude is extended to Dr. Conor T. McLennan, Dr. Sven L. Mattys, and two anonymous reviewers for their helpful comments on a previous version of this manuscript.

References

Church BA, Schacter DL. Perceptual specificity of auditory priming: Implicit memory for voice intonation and fundamental frequency. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1994; 20:521–533.

Clarke CM, Garrett MF. Rapid adaptation to foreign-accented English. Journal of the Acoustical Society of America. 2004; 116:3647–3658. [PubMed: 15658715]

Craik FIM, Tulving E. Depth of processing and the retention of words in episodic memory. Journal of Experimental Psychology: General. 1975; 104:268–294.

Delattre PC, Liberman AM, Cooper FS. Acoustic loci and transitional cues for consonants. Journal of the Acoustical Society of America. 1955; 27:769–773.

Foss DJ. Decision processes during sentence comprehension: Effects of lexical item difficulty and position upon decision times. Journal of Verbal Learning and Behavior. 1969; 8:457–462.

Goldinger SD. Words and voices: Episodic traces in spoken word identification and recognition memory. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1996; 22:1166–1183.

Goldinger SD. Echoes of echoes? An episodic theory of lexical access. Psychological Review. 1998; 105:251–279. [PubMed: 9577239]

Grossberg, S. The adaptive self-organization of serial order in behavior: Speech, language, and motor control. In: Schwab, EC.; Nusbaum, HC., editors. Pattern recognition by humans and machines (Vol. 1): Speech perception. New York: Academic Press; 1986. p. 187-294.

Hillenbrand J, Getty LA, Clark MJ, Wheeler K. Acoustic characteristics of American English vowels. Journal of the Acoustical Society of America. 1995; 97:3099–3111. [PubMed: 7759650]

Kalikow DN, Stevens KN, Elliot LL. Development of a test 80 of speech intelligibility in noise using sentence materials with controlled word predictability. Journal of the Acoustical Society of America. 1977; 61:1337–1351. [PubMed: 881487]

Klatt, DH. The problem of variability in speech recognition and in models of speech perception. In: Perkell, JS.; Klatt, DH., editors. Invariance and variability in speech processes. Hillsdale, NJ: Erlbaum; 1986. p. 300-319.

Krestar ML, McLennan CT. Examining the effects of varation in emotional tone of voice on spoken word recognition. The Quarterly Journal of Experimental Psychology. 2013; 66:1793–1802. [PubMed: 23405913]

Ladefoged P, Broadbent DE. Information conveyed by vowels. Journal of the Acoustical Society of America. 1957; 29:98–104.

McLennan CT, González J. Examining talker effects in the perception of native-and foreign-accented speech. Attention, Perception, & Psychophysics. 2012; 74:824–830.

McLennan CT, Luce PA. Examining the time course of indexical specificity effects in spoken word recognition. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2005; 31:306–321.

McLennan CT, Luce PA, Charles-Luce J. Representation of lexical form. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2003; 29:539–553.

Magnuson JS, Nusbaum HC. Acoustic differences, listener expectations, and the perceptual accommodation of talker variability. Journal of Experimental Psychology: Human Perception and Performance. 2007; 33:391–409. [PubMed: 17469975]

Mattys SL, Liss JM. On building models of spoken-word recognition: When there is as much to learn from natural “oddities” as artificial normality. Perception & Psychophysics. 2008; 70:1235–1242. [PubMed: 18927006]

Miller, JL. Effects of speaking rate on segmental distinctions. In: Eimas, PD.; Miller, JL., editors. Perspectives on the study of speech. Hillsdale, NJ: Erlbaum; 1981. p. 39-74.

Mullennix JW, Pisoni DB, Martin CS. Some effects of talker variability on spoken word recognition. Journal of the Acoustical Society of America. 1989; 85:365–378. [PubMed: 2921419]

Munro MJ. The effects of noise on the intelligibility of foreign-accented speech. Studies in Second Language Acquisition. 1998; 20:139–153.

Munro MJ, Derwing TM. Foreign accent, comprehension, and intelligibility in the speech of second language learners. Language Learning. 1995; 45:73–97.

Nygaard LC, Burt SA, Queen JS. Surface form typicality and asymmetric transfer in episodic memory for spoken words. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2000; 26:1228–1244.

Nygaard LC, Sommers MS, Pisoni DB. Speech perception as a talker-contingent process. Psychological Science. 1994; 5:42–46. [PubMed: 21526138]

Palmeri TJ, Goldinger SD, Pisoni DB. Episodic encoding of voice attributes and recognition memory for spoken words. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1993; 19:309–328.

Peterson GE, Barney HL. Control methods used in a study of the vowels. Journal of the Acoustical Society of America. 1952; 24:175–184.

Schacter DL, Church BA. Auditory priming: Implicit and explicit memory for words and voices. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1992; 18:915–930.

Theodore RM, Miller JL. Characteristics of listener sensitivity to talker-specific phonetic detail. Journal of the Acoustical Society of America. 2010; 128:2090–2099. [PubMed: 20968380]

Theodore RM, Miller JL, DeSteno D. Individual talker differences in voice-onset-time: Contextual influences. Journal of the Acoustical Society of America. 2009; 125:3974–3982. [PubMed: 19507979]

Vitevitch MS, Donoso A. Processing of indexical information requires time: Evidence from change deafness. The Quarterly Journal of Experimental Psychology. 2011; 64:1484–1493. [PubMed: 21678230]

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

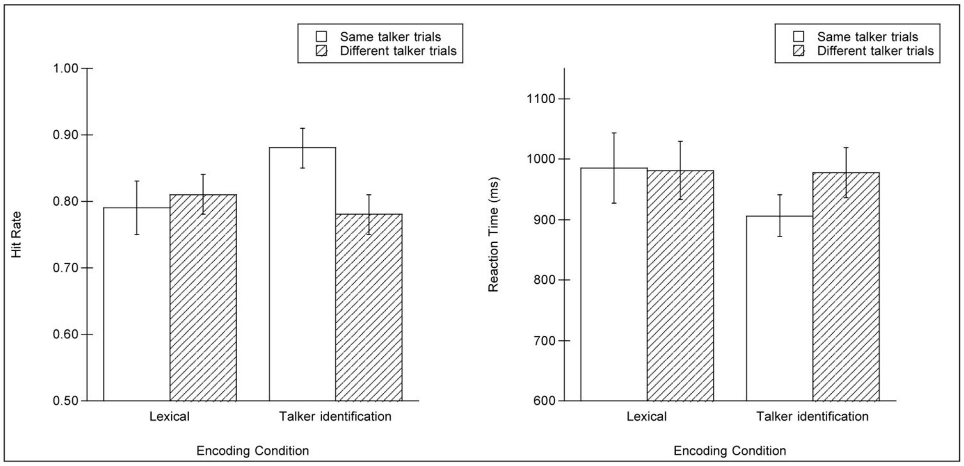

Figure 1.

Mean hit rate (left panel) and reaction time (in milliseconds, right panel) for hits during the recognition phase of Experiment 1 for each encoding condition, for same talker and different talker trials. Error bars indicate standard error of the mean.

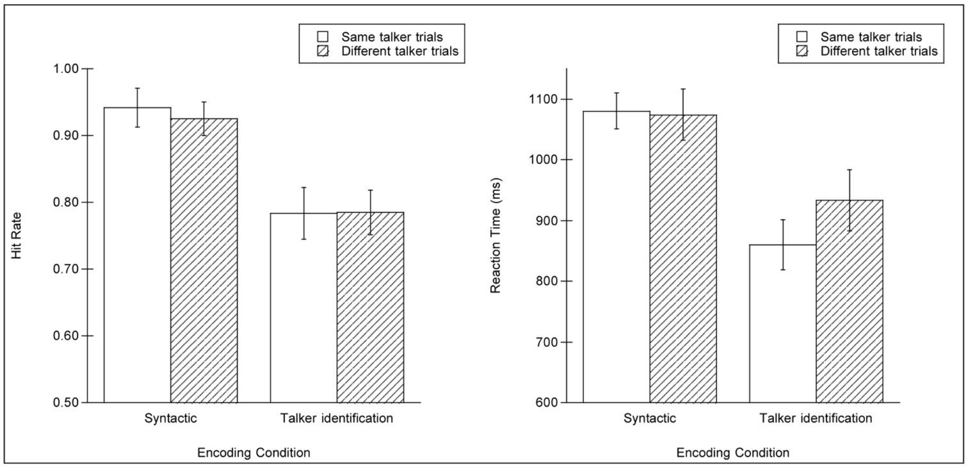

Figure 2.

Mean hit rate (left panel) and reaction time (in milliseconds, right panel) for hits during the recognition phase of Experiment 2 for each encoding condition, for same talker and different talker trials. Error bars indicate standard error of the mean.

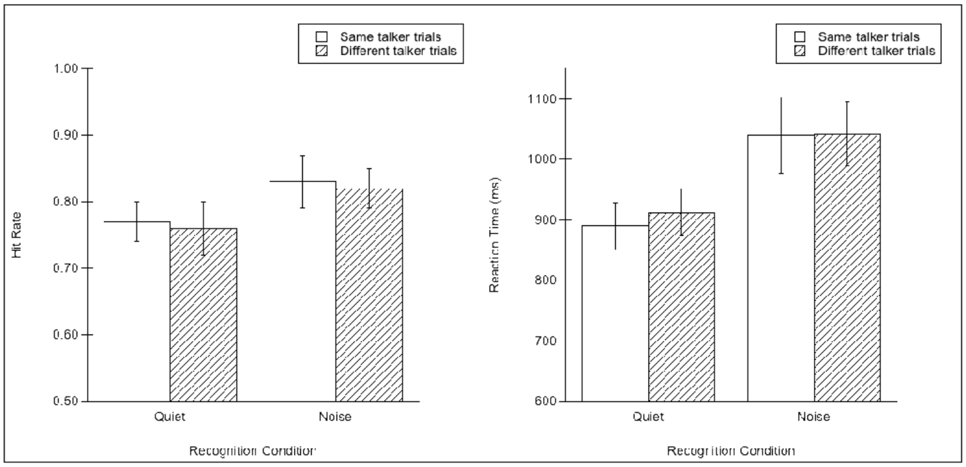

Figure 3.

Mean hit rate (left panel) and reaction time (in milliseconds, right panel) for hits during the recognition phase of Experiment 3 for each recognition condition, for same talker and different talker trials. Error bars indicate standard error of the mean.