HHS Public Access

Author manuscript

J Exp Psychol Gen. Author manuscript; available in PMC 2016 March 01.

Published in final edited form as:

J Exp Psychol Gen. 2016 March ; 145(3): 314–337. doi:10.1037/xge0000135.

Eye Movements Reveal Fast, Voice-Specific Priming

Megan H. Papesh,
Louisiana State University

Stephen D. Goldinger, and
Arizona State University

Michael C. Hout
New Mexico State University

Abstract

In spoken word perception, *voice specificity effects* are well-documented: When people hear repeated words in some task, performance is generally better when repeated items are presented in their originally heard voices, relative to changed voices. A key theoretical question about voice specificity effects concerns their time-course: Some studies suggest that episodic traces exert their influence late in lexical processing (the *time-course hypothesis*; McLennan & Luce, 2005), whereas others suggest that episodic traces influence immediate, online processing. We report two eye-tracking studies investigating the time-course of voice-specific priming within and across cognitive tasks. In Experiment 1, participants performed modified lexical decision or semantic classification to words spoken by four speakers. The tasks required participants to click a red "×" or a blue "+" located randomly within separate visual half-fields, necessitating trial-by-trial visual search with consistent half-field response mapping. After a break, participants completed a second block with new and repeated items, half spoken in changed voices. Voice effects were robust very early, appearing in saccade initiation times. Experiment 2 replicated this pattern while changing tasks across blocks, ruling out a response priming account. In the General Discussion, we address the time-course hypothesis, focusing on the challenge it presents for empirical disconfirmation, and highlighting the broad importance of indexical effects, beyond studies of priming.

Spoken word recognition is a complex process, involving the appreciation of highly variable speech signals as discrete, meaningful units. Despite superficial variations, including differences in talker identity, amplitude, speaking rate, pitch, and other idiosyncratic details (all commonly referred to as *indexical variations*; Abercrombie 1967; Pisoni, 1993), speech recognition is typically robust. Listeners fluently recognize words and discourse, with no apparent hindrance from surface variation. But what happens to the idiosyncratic details that accompany spoken words? Are they stored in memory as integral components of those words, do they fade away, or are they stored as "separate," non-phonetic information? The answers to such questions are central for evaluating theories of lexical representation and access. In fact, beyond spoken word perception, many theories in perception, memory and categorization rely critically on the general premise of "content addressable memory." In

Correspondence may be addressed to mpapesh@lsu.edu, goldinger@asu.edu, or mhout@nmsu.edu.

broad terms, there are two ways that perceptual stimuli may contact stored information in the brain. One is based on pointers, or addresses: When hearing a spoken word, seeing a face, etc., "address addressable" systems use that input to help locate the appropriate stored, abstract representation. In contrast, in content addressable memory (as in connectionist theories; Bechtel & Abrahamsen, 2002; Hinton & Anderson, 1981), information is retrieved "directly," by using stimulus features as retrieval cues to activate stored knowledge. Thus, partial information can activate perception or recall of entire stimulus assemblies. A key advantage in content-addressable systems is noise tolerance: "Address addressable" systems function by resolving noise to isolate stable cues. Content-addressable systems can use "noise" (e.g., voice-specific features) to directly access prior memories.

There are certain behavioral domains that naturally encourage content-addressable accounts. For example, in face perception, we are able to simultaneously classify faces into categories such as "man" or "woman," estimate ages, appreciate ethnic variations, etc. But we also fluently recognize friends and family, despite variations in appearances or contexts. Therefore, theories of face perception are often geared toward explaining how variable inputs activate specific stored knowledge (e.g., Lewis, 2004; Valentine, 1991), allowing us to recognize personal acquaintances. Other domains, however, encourage more abstract accounts. For example, reading is naturally characterized as a combinatorial process, wherein (in English) 26 letters are recombined into thousands of words. As such, theoretical accounts tend to treat reading as a constructive process, such that small units are rapidly combined to "unlock" stored knowledge. The foregoing contrast, between pointer systems and content-addressable memory, has long characterized a theoretical debate in speech perception. Indeed, the central theoretical question in speech perception is how listeners cope with multiple, overlapping cues to segment and word identity (e.g., Apfelbaum, Bullock-Rest, Rhone, Jongman & McMurray, 2014). The idiosyncratic signal changes introduced by different speakers can be viewed either as adding yet more noise and complexity to speech processing (e.g., Neary, 1997), or they can be viewed as beneficial information that constrains phonetic interpretations (e.g., McMurray & Jongman, 2011; Smits, 2001). The adjudication between these divergent views has long been one of main goals in theories of speech (and spoken word) perception.

By necessity, all theories of word perception posit some form of abstraction, suggesting that speech input activates sublexical units, codes that allow content-addressable access to lexical knowledge (Luce & Pisoni, 1998; Norris, 1994; Stevens, 2002). By most accounts, although indexical variations are clearly noticed (and used) by the listener, once those details have been exploited for speech decoding, they have no linguistic role and are likely forgotten (e.g., Lahiri & Marslen-Wilson, 1991; Marslen-Wilson & Warren, 1994). Abundant evidence (and common intuition) supports this view: After a conversation, people typically remember the messages that were shared, with relatively little memory for precise wording or sound patterns. By its very nature, language requires a person to fluently recombine small sets of speech units (e.g., segments, syllables, words) into new, meaningful strings – such behavior requires abstract representations at multiple levels (e.g., Chomsky, 1995). At the same time, however, people learn the vocal habits of their friends, speech imitation is common, and perceptual learning occurs for accented or idiosyncratic speech.

Intuitively, both linguistic and indexical information affect the experience of language. We hear messages, and we hear people. Clearly, the linguistic dimension is primary: It carries far more information (in bits per second), varies far more quickly over time, and constitutes the reason for speaking in the first place. The indexical dimension is inherently less important, and provides relatively little novel information from moment to moment in normal speech (although emotional tones may change relatively quickly). Nevertheless, despite their asymmetric psychological importance, both dimensions of speech are interwoven into the signal, and neither can be attended in the absence of the other. For this reason, some theories suggest that words and voices (broadly speaking) are encoded in memory as holistic exemplar representations, with the potential to facilitate later processing of perceptually similar words (e.g., Goldinger, 1996). Although such exemplar views are fairly extreme, other models propose that voice-specific information helps constrain immediate phonetic processing. For example, in the C-CuRE (Computing Cues Relative to Expectations) model, speech perception entails the simultaneous coding of multiple, overlapping spectral cues (Apfelbaum et al., 2014; McMurray & Jongman, 2011). Although some raw cues are reliable, C-CuRE optimizes processing by computing cue values in relative terms, such as appreciating that an F0 cue is high for one speaker, but low for another, and changing the evidence for voicing accordingly. Thus, voice-specific information is critically involved in "abstract" speech classification, from the earliest moments of perception. As the level of general principles, a theory such as C-CuRE offers an approach to content-addressable memory, and could (for example) be used to explain how a familiar melody can be recognized when played on a strange instrument.

Empirically, it is well-known that speaker variability across words creates processing "costs" for listeners. These are observed in perception[1] in young adults (Mullennix, Pisoni, & Martin 1989; Magnuson & Nusbaum, 2007; Nusbaum & Morin, 1992), hearing-impaired adults (Kirk, Pisoni, & Miyamoto, 1997), elderly adults (Sommers, 1996) and preschool children (Ryalls & Pisoni, 1997). Similar costs are observed in word learning: Explicit memory is reduced for word lists presented in multiple voices (Goldinger, Pisoni, & Logan, 1991; Martin, Mullennix, Pisoni, & Summers, 1989). The evidence suggests that attention is captured by abrupt speaker changes, perhaps because they force the listener to "recalibrate" their internal models for deriving the segmental content of words (Apfelbaum et al., 2014; Mullennix et al., 1989). At the same time, word-to-word speaker variations are also encoded into memory. Word perception typically improves with repeated presentations (the *repetition effect*; Jacoby & Brooks, 1984; Jacoby & Dallas, 1981; Jacoby & Hayman, 1987). When idiosyncratic surface details such as voice are preserved, this facilitation is enhanced (e.g., Church & Schacter, 1994; Schacter & Church, 1992).[2]

Across many studies, different-voice repetitions yield performance decrements, relative to same-speaker repetitions (Bradlow, Nygaard, & Pisoni, 1999; Craik & Kirsner, 1974; Fujimoto, 2003; Goh, 2005; Goldinger, 1996, 1998; Palmeri et al., 1993; Sheffert, 1998a;

[1]Note, however, that young children (infants through preschool) benefit from high-talker-variability in word learning (Richtmeier, Gerken, Goffman, & Hogan, 2009; Rost & McMurray, 2009), and that adults learning a second language also benefit from high variability (Bradlow et al., 1996; Cloppers & Pisoni, 2004; Lively et al., 1993).

[2]As reviewed by Goldinger (1996), such specificity effects are quite common. Although we focus on voice-specific priming, similar effects arise in printed word perception (font-specific priming) and music perception (e.g., Creel & Tumlin, 2012).

1998b; Yonan & Sommers, 2000). Such *voice specificity effects* arise in both perceptual and memorial tasks. For example, Goldinger (1996) found long-lasting voice effects in explicit and implicit memory. And, by measuring perceptual similarity (using MDS) among the voices, Goldinger found that priming was highly specific: As the similarity between the study and test voices increased, so did recognition accuracy (see also Goh, 2005).

Beyond "off-line" memory measures, voice effects are occasionally (but not always) observed in perceptual RTs (e.g., Meehan & Pilotti, 1996), suggesting that episodic traces of encoded words affect the online perception of test words. For example, Goldinger (1998) found that shadowing RTs were shorter for same-voice (SV) word repetitions, relative to different-voice (DV) repetitions. Moreover, while shadowing, participants spontaneously imitated words they had previously encountered, even if the last encounter was several days earlier (Goldinger & Azuma, 2004), a finding that suggests great fidelity in voice-specific memory for prior tokens. These results were modeled with MINERVA 2 (Hintzman, 1986, 1988), an exemplar model in which each encountered word creates a new memory trace. The model successfully predicted both shadowing RTs and degrees of imitation, and was sensitive to word frequency and recency, as were participants. Episodic (or exemplar) theories of word perception have proliferated (e.g., Johnson, 2006; Pierrehumbert, 2001; Walsh, Möbius, Wade, & Schütze, 2010) because – like exemplar models of perceptual classification – they are able to explain "abstract" behavior while retaining sensitivity to specific experiences. The same dual benefits arise in relative coding models, such as C-CuRE (McMurray & Jongman, 2011), which derive abstract codes by using specific, voice-level cues. (Although C-CuRE has not been tested as a model for specificity effects, it appears to be relatively straightforward extension.) Although assumptions differ across models, they generally all assume that segmental, semantic, and indexical information are encoded together. Presenting that same stimulus complex later will (potentially) activate its prior memory trace, affecting performance (although, as noted by Orfanidou, Davis, Ford & Marslen-Wilson, 2011, the locus of such effects may be response-based, rather than perceptual).

The Time-Course Hypothesis

Based on voice specificity effects, many theories of word perception now assume that both abstract and episodic information can affect processing. A theoretically critical question, however, regards the *time-course* of specificity effects in the flow of lexical processing. The question is whether episodic memory traces affect early perceptual stages, or whether abstract representations truly "drive" perceptual processing. (In this context, it is important to note that all models of word perception – whether episodic or abstract – can easily incorporate structures such as phonemes, which could theoretically dominate early processing; e.g., Wade et al., 2010.) McLennan, Luce, and colleagues suggest that voice specificity effects arise late in processing, after abstract sublexical and lexical units have already been perceptually resolved, or nearly resolved (Luce & Lyons, 1998; Krestar & McLennan, 2013; McLennan & Luce, 2005), a prediction called the *time-course hypothesis*. Conversely, others have suggested that episodes influence the earliest moments of perception (e.g., Creel, Aslin, & Tanenhaus, 2008; Goldinger, 1998).

To examine the time-course hypothesis, McLennan and Luce (2005) manipulated the relative ease of differentiation in a lexical decision task (by manipulating the phonotactic probabilities of the nonwords), and response delays in a shadowing task. They predicted that difficult lexical decisions and delayed shadowing would yield stronger specificity effects, owing to extended processing times. Indeed, when lexical decisions were difficult, voice specificity effects emerged. When the task was easier, only word repetition (priming) effects were observed. Additionally, voice effects only emerged in delayed shadowing RTs. This is opposite from the pattern observed by Goldinger (1998), although for several methodological reasons, it is difficult to compare across studies, which involved different experimental tasks, and included different numbers of words and voices. Further, one indexical variation used by McLennan and Luce (2005) was a manipulation of speaking rate, which may interact with voice-specific priming (see General Discussion). Given these methodological differences, the studies cannot be directly compared, making it difficult to appreciate the time-course of voice specificity effects.

A strength of the time-course hypothesis is its basis in a well-defined speech processing model, adaptive resonance theory (ART, Grossberg, 1980; 1999; 2003). According to ART (specifically, variants called ARTPHONE and ARTWORD; Grossberg, Boardman & Cohen, 1997; Grossberg & Myers, 2000), conscious speech perception is an emergent property of resonant feed-forward/feedback loops, acting upon speech units of all grain sizes. In brief, processing in ART spreads upward, as smaller perceptual units activate larger units, with feedback loops between levels. Initially, feature input activates *items* in working memory, which then activate *list chunks* from long-term memory. List chunks reflect prior learning, and correspond to any number of feature combinations (e.g., phonemes, syllables, words). Once list chunks are activated, items continue to feed activation upward via synaptic connections, and input-consistent chunks return activation in a self-perpetuating feedback loop (a *resonance*). This resonance binds sensory input into a coherent gestalt, allowing attention to be directed to the percept and an episodic memory trace to be formed (for reviews of ART, as it pertains to speech perception, see Goldinger & Azuma, 2003; Grossberg, 2003).

Of particular relevance, according to ART, high-frequency items in memory establish resonance quickly and efficiently when activated, relative to lower-frequency items. McLennan and Luce (2005) cited this ART principle to predict late-arriving specificity effects. By their account, frequently encountered items (e.g., phonemes, biphones, syllables, common words) are *functionally abstract*, whereas combinations of those items and indexical features are not. By definition, abstract features are far more common than any idiosyncratic variations (e.g., you hear the word “waffle” more often than you hear “waffle” in any particular voice). Therefore, sublexical and lexical information will achieve resonance sooner than voice-specific information. If voice specificity effects arise, it should be when performance is relatively slow, consistent with McLennan and Luce’s (2005) findings.

Although ART predicts that high-frequency elements will resonate more quickly than rare elements, voice specificity effects typically arise after recent presentation. (This is true for specificity effects that arise in reading, music perception, category learning, and other

domains.) In many models, frequency and recency are nearly isomorphic, and repetition effects often overpower frequency effects (e.g., Scarborough, Cortese, & Scarborough, 1977). Indeed, low-frequency words elicit more repetition priming, relative to high-frequency words, and also larger voice specificity effects (Goldinger, 1998). In order for ART to explain priming, word perception must create memory traces, capable of enhancing resonance upon repetition. Thus, it is not entirely clear whether ART actually predicts the time-course hypothesis for recently encountered tokens: The recency of the whole stimulus (voice included) may overpower the frequency of its component parts, depending upon parameter settings in the model. As it happens, the empirical literature is somewhat mixed as well: McLennan (2007, p. 68) noted that "few studies offer support for the involvement of episodic representations during the immediate on-line perception of spoken language (e.g., by reporting reaction times)." Shortly thereafter, Creel, Aslin, and Tanenhaus (2008) observed early voice-specificity effects using the *visual world paradigm*.

Eye movements are an excellent medium for examining the time course of lexical access, and many studies show a tight correspondence between speech processing and eye movements. For example, Allopenna, Magnuson, and Tanenhaus (1998) found that eye fixations reveal phonological competition during spoken word comprehension. Similarly, Dahan et al. (2001) observed word frequency effects within the first moments of perception; when high- and low-frequency cohort competitors compete with a target, eye movements are preferentially drawn to the high-frequency competitor (see also Dahan & Tanenhaus, 2004; Magnuson, Tanenhaus, Aslin, & Dahan, 2003, for related findings). Creel et al. (2008) observed voice specificity effects within a few hundred milliseconds of target word onset. In their first experiment, participants viewed four pictures on a computer screen, two of which were onset competitors (e.g., *cows* and *couch*). Competitors were repeated 20 times, and were either produced by the same speaker (e.g., female-*cows* and female-*couch*) or by different speakers (e.g., female-*cows* and male-*couch*). After repeated exposures, participants' initial eye movements to the competitor object decreased in different-voice pairs, relative to same-voice pairs, suggesting that participants encoded the voices and words, facilitating later perception. This effect emerged early in processing, before word offset. In a second experiment, Creel et al. replicated this pattern for novel words, although with a slightly longer time-course.

The Present Study

Although the results from Creel et al. (2008) appear to contradict the time-course hypothesis, there are methodological factors that complicate interpretation. Perhaps most salient, the sheer number of word-voice repetitions used by Creel et al. may have "forced" their observed effect. In the visual world paradigm, icons are presented approximately 500 ms prior to speech onset: Given such intense training, participants may have learned to anticipate voices for certain images, allowing eye-movements to be triggered by raw acoustic cues, such as pitch or timbre, irrespective of phonetic processing. (Goldinger, 1998, also observed stronger voice specificity effects when word-voice pairs had been encoded multiple times.) From a theoretical perspective, we must emphasize the "stakes" of the current question. According to numerous accounts, whether they are pure exemplar theories (e.g., Goldinger, 1998; Wade et al., 2010) or relative coding theories (e.g., Apfelbaum et al.,

2014; McMurray & Jongman, 2011), the essential function of indexical information is to provide early constraints on speech processing. That is, voice information (whether activated from stored exemplars, or coded directly from the signal) shapes *ongoing perception*, leading to criterion shifts, faster processing, spontaneous imitation, etc. In contrast, the time-course hypothesis makes the directly opposite claim, such that almost all speech processing is accomplished by activating stable codes in memory, and voice-specific information is only involved as a late-arriving constraint of last resort. The evidence in favor of this latter view comes from repeated findings that voice-specific priming effects occur mainly when people respond relatively slowly (e.g., in a lexical decision task). Given its theoretically incisive nature, the time-course hypothesis requires further study.

In the present study, we examined eye-movements to assess the time-course of voice specificity effects, while reducing the disparity between the studies by McLennan & Luce (2005) and Creel et al. (2008). We developed a novel two-alternative classification method, which we applied to both lexical decision and semantic classification. The method combines spoken word classification (such as word/nonword) with simple visual search (a schematic trial is shown in Figure 1). In each trial, the participant first gazed at a central fixation cross. Once gaze was maintained for two seconds, the cross vanished and two things happened simultaneously: A spoken word (or nonword) was played over headphones, and two target objects (a red "×" and a blue "+") appeared on the screen. Depending upon condition, these symbols represented different response options, such as *word* versus *nonword*. The symbols appeared at random locations across trials, but were constrained to always appear in the same visual half-field (e.g., the red "×" would always appear on the left side of the screen), at least 8 degrees (in visual angle) from central fixation. In this manner, the task was similar to standard classification with responses mapped to left- and right-hand buttons. The participant's task was to quickly make a decision (e.g., word/nonword), and indicate that decision by finding and mouse-clicking the corresponding symbol. The task was conducted in two blocks per experiment, a "study" block that introduced the word-voice pairings and a "test" block that intermixed same- and different-voice repetitions (along with new words, used to estimate priming).

One advantage of the current method is that it afforded analysis of voice specificity effects at multiple points in time, including saccade initiation, target location with the eyes, and the eventual mouse-click. Testing saccade initiation, in particular, allowed us to assess whether voice specificity affects early word perception. In keeping with the spirit of McLennan and Luce (2005), we used low- and high-frequency words, allowing a natural division of faster and slower trials. (For reasons we address in the General Discussion, we did not manipulate task difficulty by changing the nonwords.) Manipulating word frequency also incorporated a variable that we considered likely to "work," allowing method validation, which was important given its novelty. (We also conducted two small-scale experiments, using keyboards as the response mechanism, reported in Appendix A. These experiments verify that our new procedure produces similar behavioral patterns, but with faster RTs for the time-to-initiate measure, as described below.) We report two experiments on the time-course and generality of voice-specificity effects: In Experiment 1, participants completed consecutive blocks of the same task (either lexical decision or semantic classification),

allowing assessment of within-task priming. Because such within-task priming may be response-based, rather than perceptual (Orfanidou et al., 2011), participants in Experiment 2 completed both tasks (in counterbalanced order), allowing assessment of cross-task, perceptual priming.

Experiment 1

Experiment 1 assessed the time-course of voice specificity effects in two tasks, lexical decision (LD) and semantic classification (SC). To examine the time-course, we used three different moments in an eye-tracking task, the initial saccade off of central fixation (indicating a "word/nonword" decision), fixation of the target icon, and the final mouse-click. There were several questions of interest: First was whether we could observe voice effects in the earliest measure, when participants first moved their eyes toward the left or right. Second was whether voice effects would become larger as the dependent measures were extended in time. Third was whether (at each tested interval) the observed voice effects would be related to overall RTs. The time-course hypothesis predicts that slower trials will show larger voice effects. If voice effects are mainly observed in the later-arriving dependent measures, or if they are positively correlated with RTs within measures, the time-course hypothesis would appear to be supported.[3] Alternatively, if episodic traces affect perception from its earliest moments, we should observe voice effects in the faster dependent measures, and they should not necessarily require slower processing.

Method

Participants—Ninety-three Arizona State University students participated for partial course credit; all were native English speakers with no known hearing deficits and normal, or corrected-to-normal, vision. Four participants were dropped for having excessive error rates (as described below), leaving 48 (31 female, mean age 19.6 years) participants randomly assigned to lexical decision (LD) and 41 (27 female, mean age 20.1 years) assigned to semantic classification (SC).

Stimuli—Four speakers (two male, two female) recorded a list of 220 words (110 high-frequency, HF; 110 low-frequency, LF) and 80 nonwords (NW; see Table 1 for stimulus characteristics). Among the words, 140 were selected for use in the SC condition; 70 represented items larger than a toaster, and 70 represented items smaller than a toaster (with 35 HF and LF words within each set). The remaining 80 words (and nonwords) were used in the LD condition. For stimulus recording, words were spoken at a comfortable speaking rate, and nonword pronunciation was standardized by having the speakers shadow a recording of the experimenter pronouncing the items. The average word duration was 616 ms, with no difference based on lexical variables. HF and LF words were 611 and 621 ms, respectively, $t(219) = 0.43$, ns. Nonwords were longer (708 ms), but were not compared to words in any analyses. Among the four speakers, there were small differences in average speaking rates per item: Collapsing across words and nonwords, the male speakers had

[3]As noted, the time-course hypothesis predicts that voice effects should increase as RTs increase. Although this is a valid theoretical prediction, we demonstrate in the General Discussion that such an effect is mathematically inevitable, regardless of hypothesized perceptual mechanisms.

average rates of 633 and 651 ms, respectively. The female speakers had average rates of 599 and 618 ms, respectively. In all experiments, all recorded tokens were used equally often as "same" trials, and were used equally often on "switch" trials. Thus, differences in speaking rate were controlled and would not systematically affect RTs.

Procedure—Participants were tested individually in a quiet, dimly-lit room. The experiment was conducted using an EyeLink 1000 eye-tracker, with a table-mounted camera, recording at 1000 Hz. A chin-rest was used to maintain constant distance (55 cm) to the screen. Participants were first calibrated using a 9-dot calibration routine; all were successfully calibrated within two attempts. Following calibration, the task was explained and participants completed two practice trials, hearing a voice not used in the experiment proper. Each trial began with a central fixation cross: This was a + sign, in 18-point, enclosed in an invisible interest area. The interest area was 100 × 100 pixels (with screen resolution 1024 × 768), and was 3.3 × 3.1 degrees of visual angle. The participant had to maintain gaze in this interest area for 2000 ms before a trial would continue.

After the fixation cross disappeared, participants heard a spoken item over headphones (in one of four voices, randomly selected per trial). At the same time, two target symbols appeared in randomly selected locations on the left and right sides of the screen. Based on condition, participants quickly made a semantic classification or lexical decision by locating and clicking the appropriate response option (see Figure 1). To respond "word" (or "larger than a toaster"), participants located and clicked a blue '+.' To respond "nonword" (or "smaller than a toaster"), participants located and clicked a red '×.' These symbols were each shown in 22-point font, and were enclosed in interest areas that were 2.7 × 2.2 degrees of visual angle. These interest areas were used both to record target fixations and clicks. The response icons were moved randomly on each trial, but were constrained to always appear in the same visual half-field (e.g., the red ×, for a "word" response, would always appear on the right). This regularity allowed participants to begin executing lateral eye movements as the word unfolded in time (i.e., once they had sufficient information to make a decision), but they still had to find the response option on the screen. An invisible region surrounding fixation that ensured that no response icon appeared closer than 3.5 degrees from fixation; the average distance of the response icon from fixation was 6.5 degrees.

Following the practice trials, participants received clarifying instructions (as needed) before starting the experimental trials. In LD, participants completed 120 study trials; in SC, they completed 112 study trials. In either task, all four voices were used equally often. After the study trials and a 5-minute break, LD participants completed a test block of 160 trials (120 were repeated from Block 1), and SC participants completed a test block of 142 trials (112 repeated from Block 1). Response mapping was maintained from Block 1. Of the repeated trials in Block 2, half were same-voice (SV) repetitions, whereas 25% were spoken by a new speaker of the opposite sex of the original speaker (across-gender) and 25% were spoken by a new speaker of the same sex (within-gender). Ultimately, this gender manipulation did not contribute any findings of interest, so we report findings classified simply as SV and DV. Therefore, in the second block of each task, there were new items, used to calculate priming effects, comparing them to repeated items. Among the repeated items, half were SV trials

and half were DV trials, allowing calculation of voice effects. Results from the initial encoding block were not examined. The experiment took approximately 60 minutes.

Results

Throughout this article, we focus on RTs from correct classification trials. Because our key question regarded timing of initial saccades off central fixation, we only examined trials wherein the initial saccade was in the correct lateral direction. Four participants were removed for having an excessive bias (always toward rightward movement, in > 80% of trials), with two each in the LD and SC conditions. Among the remaining participants, 94.7% of all trials were retained for RT analyses, and we do not consider accuracy further. (There were not enough errors to allow statistical analysis, with too many design cells with zero observations per participant. At a general level, however, no experiments showed any evidence of speed-accuracy trade-offs, with errors being higher in conditions with slower RTs.) Alpha was maintained at .05. Voice effects were tested as planned pairwise comparisons (separately for nonwords, high-frequency words and low-frequency words); these were treated as multiple comparisons with Bonferonni correction. All analyses were conducted both with participants and items as random factors. For clarity, the results are presented in sections, first corresponding to LD and SC. Within each of those conditions, we present analyses moving from the slowest dependent measure (mouse RTs) toward the fastest (saccade initiation).

Lexical Decision (LD)

Validation Measures: Given the nature of our questions and the novelty of our method, we first conducted several analyses to ensure the validity of the data. In particular, our goal was to assess voice specificity effects at three points in time, measured by *time to initiate* saccades (TTI), *time to fixate* the target (TTF), and *time to click* with the mouse (RT). Analytically, this involved separate ANOVAs for each dependent measure, which is conservative but cannot reveal whether the measures relate to each other: Does faster saccade initiation reliably predict faster target location or clicking? To ensure that the results were self-consistent, we first tested correlations within trials, finding high positive values for all measures: TTI was correlated with TTF (r = .78) and with RT (r = .63), and TTF was correlated with RT (r = .87; all $p < .0001$). These correlations verify that the results were orderly, and in particular that the TTI measure truly relates to overall performance. Similar results were obtained for semantic classification and in Experiment 2 (all $r > .65$), but are not reported.

Mouse RTs: The “slowest” index of processing was RT, operationalized by the time it took participants to click the response option. Table 2 shows all the LD results from Experiment 1, with the upper section showing mouse RTs for HF words, LF words, and nonwords. Each is shown as a function of old/new status and repetition voice (for repeated items). Derived *priming effects* (relative to new words) and *voice effects* (DV minus SV) are shown for each measure; asterisks indicate reliable effects. Mouse RTs were first analyzed in a 3 (Item Type: HF/LF/NW)×3 (Voice: New, SV, DV) within-subjects, repeated-measures (RM) ANOVA. A parallel item-based ANOVA had the same factors, but Item Type was a

between-subjects variable. In each ANOVA, planned comparisons (with Bonferroni adjustments) we used to assess priming and voice effects.

There was a main effect of Item Type, $F_S(2, 46) = 55.1$, $p < .001$, $\eta^2_p = .71$; $F_I(2, 157) = 38.9$, $p < .001$, $\eta^2_p = .43$: RTs were fastest to HF words, followed by LF words and nonwords, respectively. The main effect of Voice was also reliable, $F_S(2, 46) = 19.9$, $p < .001$, $\eta^2_p = .46$; $F_I(2, 156) = 22.0$, $p < .001$, $\eta^2_p = .39$, with fastest RTs to same-voice repetitions, followed by different-voice repetitions, then new items. The interaction of Item Type × Voice was not reliable, $F_S(4, 44) = 2.2$, $p = .09$; $F_I(4, 314) = 1.9$, $p = .13$, showing that voice effects were similar for all item types. Nevertheless, because of their central role in the experiment, we still evaluated voice effects for each stimulus type. Considering first HF words, reliable priming was observed for SV repetitions ($p = .009$ by subjects, $p < .001$ by items), but not for DV repetitions (p_S and p_I both > .35). The 40-ms voice effect was not reliable, (p_S and p_I both > .11). Among the LF words, reliable priming was again observed for the SV words, ($p_S < .001$; $p_I < .001$), and also for DV words ($p_S = .003$; $p_I < .001$). Despite the numerical trend, the 52-ms voice effect was not reliable, ($p_S = .09$; $p_I = .12$). Finally, among the nonwords, reliable priming was observed for SV items, ($p_S < .001$; $p_I = .006$), and was marginal for DV items ($p_S = .05$; $p_I < .002$). The 19-ms voice effect was not reliable (p_S and p_I both > .39).

Time-to-fixate (TTF): The "intermediate" processing index was participants' average times to fixate the correct response icon, operationalized by a 100-ms (or longer) fixation within an interest area surrounding either icon. As in the mouse RTs, these results (shown in the central rows of Table 2) were first analyzed in two 3×3 RM ANOVAs (by participants and items), including planned comparisons with Bonferroni-corrected alpha. We observed a main effect of Item Type, $F_S(2, 46) = 30.6$, $p < .001$, $\eta^2_p = .51$; $F_I(2, 157) = 27.0$, $p < .001$, $\eta^2_p = .55$. TTFs were fastest to HF words, followed in order by LF words and nonwords. The main effect of Voice was also reliable, $F_S(2, 46) = 16.7$, $p < .001$, $\eta^2_p = .42$; $F_I(2, 157) = 21.6$, $p < .001$, $\eta^2_p = .45$, with fastest TTFs to SV repetitions, followed by DV repetitions, then new items. The Item Type × Voice interaction was not reliable, $F_S(4, 44) = 1.6$, $p = .43$; $F_I(4, 314) = 1.5$, $p = .51$, as voice effects were similar for all item types.

As with the mouse RTs, planned comparisons were used to evaluate priming and voice effects. For HF words, reliable priming was observed for SV repetitions, ($p_S < .001$; $p_I < .001$), but not for DV repetitions, (p_S and p_I both > .61). The 41-ms voice effect was marginal ($p_S = .03$; $p_I < .06$). Among the LF words, reliable priming was again observed for the SV words ($p_S = .008$; $p_I < .001$), and also for DV words ($p_S = .002$; $p_I = .003$). The 41-ms voice effect was also reliable, ($p_S = .009$; $p_I = .002$). Finally, among the nonwords, reliable priming was observed for SV items ($p_S < .001$; $p_I < .001$), but was null for DV items (p_S and p_I both > .36). The 53-ms voice effect was reliable ($p_S = .003$; $p_I < .001$).

Time-to-initiate (TTI): In each trial, the earliest index of lexical decision was the latency to initiate saccadic eye movements (in the correct direction) off the fixation cross. As the word unfolded in time, participants could initiate a leftward or rightward saccade at any time. As in the prior RTs, these results (lower rows of Table 2) were first analyzed in two 3×3 RM ANOVAs (participants and items), followed by planned comparisons. We observed a main

effect of Item Type, $F_S(2, 46) = 35.3$, $p < .001$, $\eta^2_p = .43$; $F_I(2, 157) = 38.4$, $p < .001$, $\eta^2_p = .41$. TTIs were fastest to HF words, followed in order by LF words and nonwords. The Voice effect was also reliable, $F_S(2, 46) = 71.9$, $p < .001$, $\eta^2_p = .60$; $F_I(2, 157) = 93.1$, $p < .001$, $\eta^2_p = .58$, with fastest RTs to SV repetitions, followed by DV repetitions, then new items. The Item Type × Voice interaction was not reliable, $F_S(4, 44) = 1.8$, $p = .13$; $F_I(4, 314) = 0.9$, $p = .39$, as voice effects were similar for all item types. Among the HF words, reliable priming was observed for SV repetitions ($p_S < .001$; $p_I < .001$), but not for DV repetitions (p_S and p_I both > .14). The 46-ms voice effect was reliable ($p_S < .001$; $p_I < .001$). Among the LF words, reliable priming was again observed for the SV words ($p_S < .001$; $p_I < .001$), and also for DV words ($p_S < .001$; $p_I < .001$). The 62-ms voice effect was also reliable ($p_S < .001$; $p_I < .001$). Finally, among the nonwords, reliable priming was observed for SV items ($p_S < .001$; $p_I < .001$) and for DV items ($p_S < .001$; $p_I < .001$). The 60-ms voice effect was reliable ($p_S < .001$; $p_I < .001$).

Semantic Classification (SC)

Mouse RTs: The SC results are shown in Table 3; these were analyzed in the same manner as the LD data. The main effect of Frequency was not reliable, $F_S(1, 40) = 2.9$, $p = .09$; $F_I(1, 140) = 1.8$, $p = .25$. Mean RTs to HF and LF words were 1671 and 1647 ms, respectively. The Voice effect was reliable, $F_S(2, 39) = 37.7$, $p < .001$, $\eta^2_p = .66$; $F_I(2, 139) = 49.0$, $p < .001$, $\eta^2_p = .71$, with fastest RTs to SV repetitions, followed by DV repetitions, then new items. The Frequency × Voice interaction was also reliable, $Fs(2, 39) = 4.9$, $p < .02$, $\eta^2_p = .18$; $F_I(2, 139) = 6.6$, $p < .01$, $\eta^2_p = .23$. This interaction was driven by the new words, which showed a large (and backwards) frequency effect. In planned comparisons for the HF words, reliable priming was observed for SV repetitions ($p_S < .001$; $p_I < .001$) and for DV repetitions, ($p_S < .001$; $p_I < .001$). The 39-ms voice effect was also reliable ($p_S = .005$; $p_I = .003$). Among the LF words, reliable priming was again observed for the SV ($p_S < .001$; $p_I < .001$) and the DV repetitions ($p_S < .001$; $p_I < .001$). The 36-ms voice effect was also reliable, although it was not a strong effect, and marginal by items ($p_S = .03$; $p_I = .07$).

Time-to-fixate (TTF): The TTF results (central rows of Table 3) were analyzed in the same manner as the Mouse RT data. In the overall ANOVA, there was a main effect of Frequency, $F_S(1, 40) = 5.6$, $p = .028$, $\eta^2_p = .12$; $F_I(1, 140) = 8.1$, $p < .01$, $\eta^2_p = .35$, with faster fixations to HF words (818 ms) than to LF words (852 ms). The Voice effect was also reliable, $F_S(2, 39) = 26.6$, $p < .001$, $\eta^2_p = .40$; $F_I(2, 139) = 23.9$, $p < .001$, $\eta^2_p = .51$, with fastest RTs to SV repetitions, followed by DV repetitions, then new items. The Frequency × Voice interaction was not reliable, $F_S(2, 39) = 0.6$, $p = .54$; $F_I(2, 139) = 1.8$, $p = .28$. In planned comparisons for HF words, reliable priming was observed for SV ($p_S < .001$; $p_I < .001$) and DV repetitions ($p_S < .001$; $p_I < .001$). The 58-ms voice effect was also reliable ($p_S = .004$; $p_I = .006$). Among LF words, reliable priming was again observed for SV ($p_S < .001$; $p_I < .001$) and DV trials ($p_S < .001$; $p_I < .001$). The 70-ms voice effect was also reliable ($p_S = .007$; $p_I < .001$).

Time-to-initiate (TTI): The TTI results (lower rows of Table 3) were analyzed in the same manner as the foregoing data. In the overall ANOVA, there was no Frequency effect, $F_S(1, 40) = 0.3$, $p = .57$; $F_I(2, 139) = 0.4$, $p = .65$, with equivalent initiation times to HF (517 ms)

and LF (522 ms) words. The Voice effect was reliable, $F_S(2, 39) = 83.4$, $p < .001$, $\eta^2_p = .77$; $F_I(2, 139) = 79.0$, $p < .001$, $\eta^2_p = .61$, with fastest RTs to SV repetitions, followed by DV repetitions, then new items. The Frequency × Voice interaction was not reliable, $F_S(2, 39) = 1.7$, $p = .19$; $F_I(2, 139) = 0.9$, $p = .58$. In planned comparisons for HF words, reliable priming was observed for SV ($p_S < .001$; $p_I < .001$) and DV repetitions ($p_S < .001$; $p_I < .001$). The 37-ms voice effect was also reliable ($p_S = .002$; $p_I < .001$). Among the LF words, reliable priming was again observed for SV ($p_S < .001$; $p_I < .001$) and DV trials ($p_S < .001$; $p_I < .001$). The 39-ms voice effect was also reliable ($p_S = .008$; $p_I = .004$).

Correlations: As noted earlier, a key prediction from the time-course hypothesis is that, as responses become slower, the strength of voice-specificity effects should increase. We tested this prediction for all dependent measures, in both LD and SC, analyzing words (excluding nonwords from LD), collapsed across participants. Before turning to the results, we must note two important points. First, although this prediction is inherent to the time-course hypothesis, prior studies (e.g., McLennan & Luce, 2005; Krestar & McLennan, 2013) have divided items into "easier" and "harder" groups and examined results categorically, rather than continuously. Our inclusion of these analyses is meant to illustrate a point that is elaborated in the General Discussion. Specifically, a challenge arises for the time-course hypothesis, as its central prediction (larger specificity effects given slower RTs) may reflect statistical properties of RTs, rather than any underlying psychological principle. Second, when examining Tables 2 and 3, it is clear that robust frequency effects were observed, and that priming effects (old versus new) were generally stronger for LF words, relative to HF words. (The sole exception to this pattern was mouse-click RTs in SC.) Nevertheless, voice-specificity effects were nearly equivalent for LF and HF words, with an average difference of only 6.5 ms. Thus, although category-level data did not follow the time-course hypothesis prediction, it remained possible that the trial-level data would show such an effect.

To test whether voice-specificity effects were indeed larger for more challenging items, we calculated mean RTs for each repeated word, separately for those trials that were SV and DV repetitions.[4] The average RT per word was computed (averaging SV and DV trials); voice effects were calculated as the difference score, DV minus SV. In every case, the correlations of mean RTs and voice effects were positive and reliable: In LD, the correlations for mouse RTs, TTF, and TTI were $r = .19$ ($p < .05$), $r = .32$ ($p < .001$) and $r = .37$ ($p < .001$), respectively. In SC, these same correlations were $r = .16$ ($p < .05$), $r = .19$ ($p < .05$) and $r = .30$ ($p < .001$). Although the pattern cannot be appreciated by grouping words into LF and HF categories, there were reliable associations between item RTs and voice-specificity effects.

Discussion

In Experiment 1, we created a modified 2AFC method, applied to both lexical decision and semantic classification. The method uses lateral eye-movements to provide an early measure of classification, with later measures of target location and selection. At a broad level, the

[4]Because we used four voices, this approach still combined three different tokens per word into the "DV" category. Testing tokens separately did not substantially change the results reported here.

results validated the method. By all three dependent measures, the results were orderly and expected: Word frequency, repetition priming, and voice specificity effects were observed at all time-points in both LD and SC (although voice effects were not reliable in mouse-click RTs in LD, there were consistent numerical trends).

With respect to the time-course hypothesis, Experiment 1 provided mixed evidence. Applied to the present method, the hypothesis most comfortably predicts that voice effects should be stronger in later-arriving measures of lexical processing, such as mouse-click RTs. Nevertheless, we observed robust voice effects in saccade initiation times; these were replicated in the later measures. Indeed, given repeated words in LD, people initiated saccades in the correct direction an average of 12 ms after word offset; this value was 68 ms in SC. (These results are reminiscent of those from Creel et al., 2008, who observed voice-specific effects on eye-movements prior to word offset.) On the other hand, by all dependent measures, we also observed positive associations: As response times grew slower, voice-specificity effects grew larger, as the time-course hypothesis would predict.

Taken together, the results from Experiment 1 show that voice-specificity effects can arise early in lexical processing, although those effects positively correlated with overall RTs to different words. A key question, however, is the degree to which the effects in Experiment 1 may have reflected response priming, rather than perceptual priming. In SV trials, participants were required to make the same response (e.g., *word-nonword*) to the same recorded token. Strong repetition effects were observed, but cannot be clearly interpreted as perceptual effects (see Orfanidou et al., 2011). Instead, repetitions could serve as episodic memory probes, reactivating recent responses made to those same items. Although such an interpretation may be interesting in its own right, our present goal was to assess perceptual specificity effects. In Experiment 2, we crossed the study and test procedures from Experiment 1, such that half the participants completed LD first, followed by SC; the other half completed both tasks in the opposite order. The crossed procedure made it unlikely for priming to be response-driven, allowing greater focus on perceptual processes.

Experiment 2

The goal of Experiment 2 was to determine whether voice-specific priming would occur across changes in task. Participants again completed an initial blocks of trials, either LD or SC, then completed the complementary task in a second block.

Method

Participants—Fifty native English speakers (31 female, mean age 19.9 years), with no self-reported hearing deficits and normal, or corrected-to-normal, vision participated in exchange for partial course credit. Half the participants were randomly assigned to complete LD first, followed by SC; the other half completed tasks in the reverse order. Two participants were dropped from analysis for having excessive errors (as in Experiment 1, both showed consistent rightward bias > 85% of trials), leaving 48 in the final analysis.

Stimuli—The materials from Experiment 1 were used in Experiment 2.

Procedure—The procedures were identical to Experiment 1, except all participants completed both tasks in counterbalanced order. Half of the participants first completed 160 LD trials (80 words and 80 nonwords). After a short break, they completed 112 SC trials, including 80 words that had been presented in Block 1. The remaining participants first completed a block of 64 SC trials, followed by 160 LD trials in Block 2, with 64 words repeated from Block 1. In the second block of either condition, half the repeated words retained their original presentation voices, and half were changed to one of three other voices.

Results

After removing errors, 91% of trials were retained for RT analyses; we do not consider accuracy further. Because the LD and SC blocks were presented in counterbalanced order (to avoid task-order effects), all analyses were conducted on average Block 2 data, collapsing across tasks. (For interested readers, Appendix B shows results from the different task orders. Perhaps not surprisingly, SC had greater priming effects on LD, relative to the reverse. For purposes of drawing conclusions, we limit analyses to the full design.) The results (for words only) are shown in Table 4, which is arranged in similar fashion to Tables 2 and 3. As in Experiment 1, we first conducted omnibus ANOVAs (by both participants and items), with planned comparisons for priming and voice effects.

Mouse RTs—The upper rows of Table 4 show mouse-click RTs for low- and high-frequency words, each as a function of old/new status and voice (for repeated items). Derived priming and voice effects are shown, as before. Mouse RTs were first analyzed in a 2 (Frequency: HF/LF) × 3 (Voice: New, Same, Different) RM ANOVA. There was no main effect of Frequency, $F_S(1, 47) = 0.61$, $p = .71$; $F_I(1, 79) = 1.31$, $p = .60$. The numerical trend was for a "backwards" frequency effect, with RTs 26 ms faster to LF words, relative to HF words. The Voice effect was reliable, $F_S(2, 46) = 36.8$, $p < .001$, $\eta^2_p = .62$; $F_I(2, 77) = 28.0$, $p < .001$, $\eta^2_p = .39$, with fastest RTs to SV repetitions, followed by DV repetitions, then new items. The Item Type × Voice interaction was not reliable, $F_S(1, 47) = 1.20$, $p = .49$; $F_I(2, 77) = 0.99$, $p = .58$. Among the HF words, reliable priming was observed for SV ($p_S < .001$; $p_I < .001$) and DV repetitions ($p_S = .006$; $p_I = .002$). The 11-ms voice effect not reliable. Among the LF words, reliable priming was again observed for SV ($p_S < .001$; $p_I < .001$) and DV trials ($p_S = .03$; $p_I = .009$). The 74-ms voice effect was also reliable ($p_S = .008$; $p_I = .008$).

Time-to-fixate (TTF)—The TTF results (central rows of Table 4) were analyzed in the same manner as the mouse RTs. The main effect of Frequency was again numerically backwards (by 11 ms) and unreliable. The main effect of Voice was reliable, $F_S(2, 46) = 10.9$, $p < .001$, $\eta^2_p = .32$; $F_I(2, 77) = 15.5$, $p < .001$, $\eta^2_p = .35$, with fastest RTs to SV repetitions, followed by DV repetitions, then new items. The Frequency × Voice interaction was null. For HF words, reliable priming was observed for SV ($p_S = .005$; $p_I = .001$) and DV repetitions, ($p_S = .004$; $p_I = .002$). The 33-ms voice effect was not reliable (p_S and p_I both > .12). Among the LF words, reliable priming was again observed for SV trials ($p_S = .003$; $p_I < .001$), but not for DV trials (p_S and p_I both > .09). The 44-ms voice effect was reliable ($p_S = .03$; $p_I = .007$).

Time-to-initiate (TTI)—The TTI results (lower rows of Table 4) were analyzed in the same manner as the prior results. The main effect of Frequency was reliable, with a 45-ms effect in the expected direction, $F_S(1, 47) = 16.0$, $p < .001$, $\eta^2_p = .36$; $F_I(1, 79) = 21.8$, $p < .001$, $\eta^2_p = .43$. The Voice effect was reliable, $F_S(2, 46) = 7.4$, $p < .01$, $\eta^2_p = .24$; $F_I(2, 77) = 9.1$, $p < .01$, $\eta^2_p = .30$, with fastest RTs to SV repetitions, followed by DV repetitions, then new items. The Frequency × Voice interaction was null. Among the HF words, reliable priming was observed for SV repetitions ($p_S = .007$; $p_I = .008$), but not for DV repetitions (p_S and p_I both > .29). The 46-ms voice effect was reliable ($p_S = .02$; $p_I = .006$). Among the LF words, reliable priming was again observed for the SV ($p_S = .003$; $p_I < .001$) and DV trials ($p_S = .005$; $p_I = .002$). The 38-ms voice effect was reliable ($p_S = .009$; $p_I = .004$).

Correlations—As in Experiment 1, the group-level data shown in Table 4 show little support for the time-course hypothesis: The overall voice specificity effect was 22 ms larger for LF words, relative to HF words, but that difference mainly reflected the mouse RTs. At the level of individual trials, however, we again found positive correlations between overall RTs (averaging SV and DV repetitions) and voice effects (DV minus SV). Combining all words together, the observed correlations for mouse RTs, TTF, and TTI were $r = .22$ ($p < .05$), $r = .29$ ($p < .01$) and $r = .28$ ($p < .01$), respectively. Thus, although the predicted pattern was not evident at the categorical level, there were reliable associations between item RTs and voice-specificity effects.

Discussion

Experiment 2 conceptually replicated Experiment 1, while reducing the likelihood for response priming. Notably, the response times shown in Table 4 were slower overall (by roughly 150 ms, on average) than those in Tables 2 and 3, suggesting that response priming likely did affect Experiment 1. Nevertheless, in both experiments, reliable priming occurred and voice-specificity effects showed the same general pattern: They were numerically evident in all dependent measures, and were robust in the earliest, TTI measure. Indeed, among HF words, voice effects were only reliable in the TTI data (they were evident for all measures among the LF words). Finally, as in Experiment 1, although we found little evidence for the time-course hypothesis (McLennan & Luce, 2005) at the broad level of LF versus HF words, we found reliable correlations of item RTs and voice effects. Thus, we have verified that specificity effects can arise early in perception: In Experiment 2, laterally correct eye-movements were initiated approximately 76 ms after the offsets of repeated words. There remained, however, some evidence consistent with the time-course hypothesis, which we consider further below.

General Discussion

The present study was motivated by a theoretical discrepancy in the literature, regarding the time-course of voice specificity effects in spoken word recognition. Although this sounds like a fairly nuanced topic, it has great theoretical gravity: Whereas episodic models (e.g., Goldinger, 1996; 1998) or relative-coding models (McMurray & Jongman, 2011) predict that indexical information mediates early perceptual processes, the time-course hypothesis (McLennan & Luce, 2005), derived from ART (Grossberg & Myers, 2000), predicts that

abstract sublexical elements will dominate early perception, with indexical effects only arising when processing is slow. Prior studies have lent support to both views, with evidence for early voice effects (e.g., Creel et al., 2008; Goldinger, 1998) but also evidence that such effects "arrive late" in processing (e.g., Krestar & McLennan, 2013; Luce & Lyons, 1998; McLennan & Luce, 2005). Reconciling these prior studies is challenging, however, as their methods differed in numerous regards. In this study, we examined the time-course hypothesis, using a method that allowed detection of early voice effects, while retaining a familiar 2AFC methodology.

The present method gave us several "moments" when voice-specific priming could be observed (time for the initial saccade, time to locate the object, and time to click). Although it was theoretically possible that voice effects would be absent in initial saccades, only arising in "later" measures, we did not consider this likely: In any given trial, the sequence of RTs were derived from a coherent underlying event, the recognition and classification of a spoken word. We thus considered saccade-initiation times as the critical dependent measure; the later-arriving measures served to validate that those eye-movements were meaningful.

In the visual world paradigm (e.g., Dahan et al., 2008; Magnuson et al., 2003), eye-movements are linked to fixated objects, each with their own names, so the "meanings" of eye movements can be directly inferred. In our case, quick left or right saccades had several layers of validation: First, they were in the correct direction more than 90% of the time. Second, they produced effects of classic variables (word frequency and repetition). Third, initial saccades typically predicted equivalent psychological effects later in time, such as correct mouse clicks. We are therefore confident that initial saccades reflected meaningful perceptual processes, and their implications were clear: From the earliest moments of word perception, effects such as frequency are evident (Dahan et al., 2001). For repeated words, effects of indexical repetition were observed very early, often before the spoken word was complete. This was seen in Experiment 1, which could have involved response priming (Orfanidou et al., 2011), but was also seen in Experiment 2, wherein response priming was likely minimized. In the remainder of this article, we address three further points. These include the interpretive challenge posed by the time-course hypothesis, the status of voice-specificity effects in the literature, and their implications for psycholinguistic theory.

The time-course hypothesis

In a series of articles, McLennan and colleagues have advanced the time-course hypothesis for voice-specificity effects in word perception. In summarizing their findings, McLennan and Luce (2005, p. 306) wrote, "… indexical variability affects participants' perception of spoken words only when processing is relatively slow and effortful." This theoretical hypothesis can be restated in two ways, being more or less restrictive. The more restrictive interpretation is that "voice-specific priming will not be observed when lexical processing is fluent." This strong version has been implicitly adopted in various studies: In addition to McLennan and Luce (2005), other reports have been framed in terms of null voice-specificity effects in "easy" conditions, with effects emerging only in "hard" conditions (e.g., Krestar & McLennan, 2013; McLennan & González, 2012; Vitevitch & Donoso,

2011). The present results, and other findings of early-arriving voice effects (e.g., Creel et al., 2008; Creel & Tumlin, 2011) seemingly refute the strong hypothesis: We presented relatively easy and hard words (based on frequency) for classification, and observed robust voice-specificity effects within 100 ms of word offset, equivalently for LF and HF words. In a recent article, Maibauer, Markis, Newell and McLennan (2014) also found early voice-specificity effects when famous voices were used.

Given such results, a strong version of the time-course hypothesis appears untenable. And in a scientific sense, it is difficult to defend a theoretical position that "certain conditions will create null results, and others will create positive results," because null results can arise for many unimportant reasons. The less restrictive version of the hypothesis focuses on the implied interaction, or the continuous relationship between variables: Voice-specificity effects are predicted to increase when lexical processing takes more time. This weaker version is more easily supported from an empirical standpoint, appears to have support in the prior literature, and is more scientifically tractable. Indeed, in the present data, although we observed early voice effects, we also consistently found significant, positive correlations between word RTs and derived voice effects. Such findings, however, beg a question: How convincing is the evidence for the time-course hypothesis? We suggest that the evidence is surprisingly weak, for three key reasons. First, prior studies supporting the time-course hypothesis are typically underpowered. Second, many prior studies have a critical design flaw (in our view), allowing an alternative account of their results. Third, we suggest that the key finding – larger voice effects with slower lexical processing – may be an inevitable outcome, given the nature of RT distributions. We briefly address the first and second points, then focus on the third.

Experimental power—The time-course hypothesis is well-motivated theoretically (from ART), and could be correct. Its prior empirical support, however, is not compelling. The study by McLennan and Luce (2005) was the first articulation of the time-course hypothesis, and produced its initial support. In their Experiment 2 (which bears the closest resemblance to the current study), people classified words and nonwords spoken in two voices (one male, one female) in two consecutive blocks: The first block exposed listeners to a set of words and voices; the second block had new and old words, with some old words presented in new voices. Reliable priming (studied versus new) was observed whether processing was easy (Experiment 2A) or hard (2B). However, no voice effect was observed in the easy condition (an 8-ms trend), but was observed in the hard condition (35 ms). This established the time-course pattern and has been widely cited, but the experimental details are not compelling. In particular, the experiment only included 24 stimuli, with only 12 words constituting the critical trials. Four words were control items; the remaining eight were repeated from Block 1, with four SV and four DV trials. Despite testing many participants (72 each in Experiments 2A and 2B), this provides a very limited data set, mathematically and linguistically. Mathematically, each participant only contributed eight relevant data points per experiment, divided into two distributions of four RTs each. Linguistically, the stimuli were also limited, with 12 monosyllabic words (*bear, bee, book, bowl, car, cat, deer, fly, key, leg, nail, nut*) featuring six phonological onsets. In our view, even though the results were reliable by participants and items, it is challenging to interpret such a narrow data set.

The issue of limited sampling in one study is not necessarily cause for great concern. However, McLennan and González (2012) later used the same stimulus set and reported more statistical details, allowing us to estimate their power (to detect voice specificity effects) at .5, indicating a .5 probability of detecting a true effect, should it be present. (In the present study, the observed power ranged from .87 to .99 for all tests of voice-specific priming.) As noted by Francis (2012; 2013), low-power experiments are more likely to create both false-positive and false-negative outcomes. McLennan and González (2012) also included 72 participants per test, suggesting that the low power derived from using few items per condition. These same stimulus items were also used by Krestar and McLennan (2013), with similar subdivision into conditions.

In a different example, Vitevitch and Donosio (2011) examined the time-course hypothesis using a “change deafness” paradigm. In their first experiment, listeners made lexical decisions; these were made relatively easy or hard by manipulating the nonwords. Harder nonwords slowed “word” decisions by 82 ms. Halfway through the experiment, the voice producing items was changed, from one male speaker to another. Afterward, participants were asked whether they noticed (1) “anything unusual” and (2) a speaker change. Out of 22 people in the “easy” condition, 14 (63%) detected the change in speaker, compared to 19 of 22 (86%) people in the “hard” condition. Clearly there was a difference: Five more people detected the voice change when the task was more challenging. However, two-thirds of participants in the easy condition *also* noticed the change, and the statistical difference between conditions was small and low-power, with each person providing a single data point. In a second experiment, using a confidence-scale method, the observed power for a similar effect was approximately .5, which is again quite low.

An alternative explanation—As noted earlier, in the current experiments, we opted not to manipulate lexical decision difficulty by changing the nonword foils. This decision was partly motivated by a desire to maintain parallel structure between the LD and SC tasks. Our deeper motivation, however, was to avoid the typical approach taken in prior experiments (e.g., Krestar & McLennan, 2013; McLennan & González, 2012; McLennan & Luce, 2005; Vitevitch & Donoso, 2011). The standard approach (making nonwords more or less word-like) invites an alternative explanation of the results. We focus on McLennan and Luce (2005) as a relevant example: As noted, listeners were assigned to groups for either easy or hard LD. Words were constant across conditions, while nonword foils were changed, allowing comparison of the same words under more or less fluent conditions. In the easy condition, the nonwords were very low-probability phonotactic sequences. For example, people had to discriminate *BACON* from *THUSHTHUDGE*. By contrast, in the hard condition, the nonwords differed from target words by one phoneme in final position, such as *BACON* versus *BACOV*. Another experiment was similar, but with monosyllables. By virtue of making nonwords more word-like in the hard condition, McLennan and Luce (2005, page 312, Table 4) increased “word” RTs by an average of 33 ms, and also observed voice-specific priming. This finding motivated the time-course hypothesis that slower lexical processing gave voice-specific features greater opportunity to affect the resonant dynamics of lexical access.

Although the time-course account is consistent with the data, does it provide the most likely interpretation? Do 33 extra milliseconds of processing truly change lexical dynamics to such a degree that voice effects will emerge? We suggest a more parsimonious interpretation. Specifically, changing the nonword foils does not merely increase average RTs. It also dramatically changes how attention must be allocated to the bottom-up speech signal. Given unusual nonwords such as *THUSHTHUDGE*, people can likely perform lexical decision with minimal attention to the signal -- the words should essentially "pop out." But in the hard condition, words and nonwords can only be discriminated by carefully listening for the final segment, and the listener cannot predict when that segment will arrive as the signal unfolds in time. Changing the nonword foils does not merely slow down processing; it forces participants to attend to fine-grained details in the speech signal. In previous studies, voice-specific priming is most robust when listeners focus on "surface" features of words, such as judging how clearly words are enunciated (Goldinger, 1996; Schacter & Church, 1992).

According to the time-course hypothesis, the challenging lexical decision task slows down processing, and that extra processing time allows voice effects to emerge. We suggest that changing the nonwords changes the listener's focus of attention, forcing careful bottom-up monitoring. This simultaneously creates episodic memory traces that more strongly represent superficial (e.g., voice) details of the signal. Note that this account is also entirely consistent with classic exemplar models (e.g., the *attention hypothesis* in Logan, 1988).

What is the correct null hypothesis?—Setting prior studies aside, the present experiments did produce a pattern predicted from the time-course hypothesis. In all measures, as mean RTs for repeated words increased, the magnitudes of voice effects also increased. This follows directly from the less restrictive version of the hypothesis, as noted above, and the present experiments had high observed power (always > .85). Nevertheless, a difficult question arises: What relationship between the variables *should* be expected, under the null hypothesis? RTs have a natural lower "boundary," and are typically distributed in a roughly ex-Gaussian fashion (Balota & Spieler, 1999; Ratcliff & Rouder, 1998). As a result, creating faster RTs (for example, by repetition priming) is inherently nonlinear: Words that typically elicit fast RTs can only be slightly improved by priming, whereas words that typically elicit slow RTs can show large benefits. By extension to the current study, any effect (such as voice-specific priming) that is compared across relatively fast and slow items will typically have larger effects on the slower end of the spectrum. This has long been recognized as the problem of *response time scaling*: As a general rule, effects tend to increase when RTs increase.

To better understand the correct null hypothesis for these experiments, we conducted a series of simulations, beginning with simple scenarios, then moving toward more faithful representations of real RT experiments. In the first four simulations (see Figure 2), we generated 10,000 artificial RTs for "same-voice" and "different-voice" repetitions of words, using different distributional assumptions. For each, we then calculated RTs and "voice effects" (DV minus SV). The upper row of Figure 2 (panels A and B) shows the results from shifted, flat distributions, with no relation between the mean and variance: SV word RTs were generated as randomly selected values between 500 and 800 ms; DV words were

randomly selected values between 575 and 875 ms. Panel A shows the frequency distributions for observed RTs, and panel B shows no relationship between mean RT and voice effect. Of course, natural RT distributions are not flat, and their standard deviations typically increase (approximately linearly) with mean increases (Wagenmakers & Brown, 2007). To better approximate natural RTs, we next simultaneously changed the means and standard deviations, and tested different underlying distributions.

In the second row of Figure 2 (panels C and D), we again selected from flat distributions, but SV words ranged from 500 to 800 ms, whereas DV words ranged from 550 to 950 ms. As shown in panel D, a positive relationship emerged between item RT and voice effects (across simulations, the correlation ranged from .25 to .31). In the third row (panels E and F), we sampled from Gaussian distributions: SV words had M = 500 and SD = 75; DV words had M = 600 and SD = 100, and the same relationship emerged. Finally, the bottom row of Figure 2 (panels G and H) show RTs sampled from Weibull distributions (SV mean = 850, DV mean = 1000, with a 100-ms increase in standard deviation). Weibull distributions have the same general shape as ex-Gaussian distributions and therefore resemble typical RTs (Balota & Spieler, 1999). As shown, the same relationship between item mean and voice effect was observed again.

We next conducted tests with more reasonable core assumptions. In the simulations shown in Figure 2, there was no relationship between the SV and DV values generated for each word; they were independent samples. Presumably, real words have some “intrinsic RT” that can be adjusted upward or downward by manipulations, such as repetition priming. To further test the observed relationship, we generated 10,000 normally-distributed RTs for SV words (M = 500, SD = 75) and created DV versions by adding normally distributed values (M = 50, SD = 100). In this manner, DV and SV values were tethered to each other, but were adjusted by random sampling, and could produce occasional “backwards” priming effects. Not surprisingly, the final SV and DV values were highly correlated (around r = .75). What is more surprising is that item RTs were highly correlated with voice effects, with typical values around r = .55 (see Figure 3). When this simulation approach was repeated using Weibull distributions, the correlation was approximately r = .35. This relationship is not affected by “normalizing” transformations, such as log- or z-transforming the data. Not surprisingly, the relationship can be reduced by Vincentizing, such as replacing all scores above two standard deviations from the mean with the cutoff score.[5] This has the effect of selectively suppressing the slowest RTs, which reduces the observed correlation.

Finally, to better simulate priming experiments, we created a distribution of word RTs (M = 600, SD = 75), then generated “primed” versions by subtracting normally distributed values, assuming that SV trials produce more priming, with less variance. For example, one simulation had SV priming with M = 100 and SD = 50; DV priming was M = 50 and SD =

[5]Note that these simulated data are not amenable to analyses that partition RT distributions into parameters such as mu, sigma, and tau (e.g., Balota & Spieler, 1999). Although such analyses would technically “work” on the data, their interpretation would not be clear. We intentionally created the SV and DV distributions with shifted mean values (mu), leaving sigma and tau as the parameters free to vary. Therefore, the results of any distributional analysis are predetermined. The observation of interest, however, is that even with a constant shift in mu, the typical distributional properties of RTs give rise to the RT scaling pattern, which is isomorphic to the time-course hypothesis.

100. As one would expect, over many trials the mean "voice effect" always approaches the selected value (e.g., 50), and the SV and DV trials are highly correlated (around $r = .8$), as one value was derived from the other. But of greater interest, the same relationship emerges: As item RTs increase, voice effects increase, with correlations ranging from .35–.45 across simulations, closely resembling the correlations observed in the present experiments.

These results are (at least initially) surprising, because there was no specified relationship between an item's randomly chosen "base RT" and the adjustments enacted to simulate priming. Every instance of SV or DV priming was identically sampled, with means that perfectly match their expected values. In this case, mean RTs were 501 and 550 ms, respectively. The standard deviations, however, were 90 and 125 ms, and this asymmetry drives the observed relationship. For any given item, the DV trial is more likely to be "extra slow," which simultaneously raises the mean RT and increases the voice effect. Given the well-known relationship between RT means and standard deviations (Wagenmakers & Brown, 2007), these simulations suggest that, irrespective of underlying lexical processes, voice effects will typically increase when processing is slower, as a function of RT sampling. This illustrates what Loftus (1978) termed "removable" interactions: Many interactions in psychological research are potentially spurious, arising from nonlinear underlying scales (see Faust, Balota, Spieler & Ferraro,1999; Wagenmakers, Krypotos, Criss & Iverson, 2012). As a relevant example, Hutchinson et al. (2008) found that semantic priming increased monotonically with the "base RTs" for different words. The theoretical motivation for the time-course hypothesis (McLennan & Luce, 2005) is well-conceived and could be entirely correct, but its supporting evidence is exactly what would be expected under the proper null hypothesis.

What is the status of voice-specificity effects?

There is a curious sociological phenomenon that arises with respect to exemplar theories in cognitive science, whether focused on perception, categorization, or memory. Specifically, common intuition verifies that every experienced moment is *simultaneously* abstract and specific: When conversing with a friend, your perception is inescapably categorical (your friend is a man, sitting at a table, wearing a blue shirt, etc.), revealing the abstraction that occurs constantly in cognition. At the same time, your perception is undeniably specific (your friend is Tony, who seems more tan than usual). It has long been understood that successful theories must simultaneously explain the generality and specificity of memory (see McClelland & Rumelhart, 1985; Underwood, 1969), and that abstract representations (e.g., prototypes) must derive from experiences with specific examples. It is also accepted that abstract representations are not all equivalent – some (e.g., high-frequency words) have "privileged" status, disparities that originate from differences in particular experiences. In these regards, abstract and specific representations are immutably interconnected. They are simultaneously experienced and mutually reinforcing: People "hear" physically absent pauses between words because linguistic knowledge imposes abstract structure on speech signals (Grossberg & Myers, 2000; McMurray & Jongman, 2011). Words are accessed more fluently when they are generally common, or when they are expected from a specific conversational partner. Episodic knowledge changes abstract processing.

Despite their near-isomorphism, abstract and episodic representations are generally treated as "competing ideas," rather than as complementary parts of a cognitive system that requires both stability and flexibility. It is self-evident that speech perception and production require abstract lexical and segmental representations. In modeling, abstract representations are tractable, and have ecological validity. Abstract representations dominate linguistic processing and memory: Messages are derived and retained in conversation, whereas superficial details are often forgotten. These facts often lead researchers to dismiss the role of indexical effects as extra-linguistic or superfluous (e.g., Bowers, 2000; González & McLennan, 2007; Mitterer & Ernestus, 2008; Pallier, Colome, & Sebastian-Galles, 2001). We consider this a missed opportunity. Although abstract representations may dominate linguistic processing, effects such as voice-specific priming help reveal how language, attention, and memory work together.

At this point, there have been numerous published articles examining indexical effects in language processing. Many are focused on voice-specific repetition priming, like the present study. Some examine explicit and implicit memory for clear and degraded words (e.g., Campeanu, Craik, & Alain, 2013; Church & Schacter, 1994; Geiselman & Bellezza, 1976; 1977; Goh, 2005; Goldinger, 1996; Pilotti et al., 2000; Pilotti & Beyer, 2002; Pilotti, Meade, & Gallo, 2003; Schacter & Church, 1992; Sheffert, 1998a; 1998b), or continuous recognition memory for words and voices (e.g., Bradlow et al., 1999; Craik & Kirsner, 1974; Palmeri et al., 1993, Sheffert & Fowler, 1995). Others are more focused on perception itself (e.g., Creel et al., 2008; Creel & Tumlin, 2011; Goldinger & Azuma, 2003; Luce & Lyons, 1998; Meehan & Pilotti, 1996). Studies such as these – experiments with studied and later repeated words – are most typically associated with "voice-specificity effects." However, there are many effects of speaker variation beyond such paradigms.

Moment-to-moment indexical variations (in voice, speaking rate, emotional tone) are well-known to command attention and broadly affect perception and memory across different populations of listeners (e.g., Kirk et al., 1997; Magnuson & Nusbaum, 2007; Martin et al., 1989; Mullennix & Pisoni, 1990; Mullennix et al., 1989; Sommers, 1996). A growing literature on *perceptual learning* shows that familiarity with specific voices has wide-ranging effects. These include helping infants isolate words in speech (Houston & Jusczyk, 2000; 2003), improving word perception (Nygaard & Pisoni, 1998; Nygaard et al., 1994; Yonan & Sommers, 2000), improving lip-reading (Rosenblum et al., 2007), changing sublexical processing (Allen & Miller, 2004; Eisner & McQueen, 2005; Kraljic & Samuel, 2005, 2007; Smith & Hawkins, 2012; Sumner, 2011; Sumner & Samuel, 2009; Theodore, Miller & DeSteno, 2009), and improving the ability to segregate overlapping voices (Johnsrude et al., 2013; Newman & Evers, 2007). Indexical variations in speech perception have wide-ranging effects on speech production, the phenomenon of *phonetic convergence* (or *alignment*; Goldinger, 1998; Goldinger & Azuma, 2004; Namy et al., 2004; Nielsen, 2011; Pardo, 2006; Shockley et al., 2004; Smith & Hawkins, 2012). Studies have elucidated the cortical processes that bind lexical and indexical information into memory traces (e.g., Gagnepain et al., 2008).

Our purpose is not to review the large and diverse literature surrounding indexical processing in speech. We merely wish to emphasize that indexical effects are robust and

multifaceted: Testing whether they reliably emerge in one particular measure (e.g., lexical decision) is an important part of the scientific process, but we should not equate any single paradigm with "language processing," writ large. Across domains, there is overwhelming evidence that indexical information matters in language perception, learning, memory, and production. Many authors appear motivated to protect abstraction-based theories from having to accommodate indexical processes, casting them as extra-linguistic, neurally segregated from the central business of linguistic processing (e.g., Mitterer & Ernestus, 2008). We suggest that a deeper understanding of language will result from considering the entire spectrum of effects.

Returning specifically to the present study, McLennan and Luce (2005) derived the time-course hypothesis from the claim in ART (Grossberg, 1980, 1999, 2003) that high-frequency units (such as segments in speech) typically achieve quick resonance and dominate perception. One clear strength of the hypothesis is its grounding in such a well-articulated theory. It is important to appreciate, however, that ART is a powerful theoretical framework, with great flexibility in predictions. Given certain assumptions, it easily predicts voice-specificity effects, even at short time-scales. In ART, bottom-up and top-down features combine to achieve resonance, thereby creating perception and guiding attention. When experiments contain numerous voice changes, attention is repeatedly drawn to those changes (Goldinger et al., 1991; Magnuson & Nusbaum, 2007), increasing the likelihood of episodic encoding. And, if a recent memory trace is available that matches new input, ART predicts that it will drive the perceptual system toward faster resonance, relative to tokens without matching traces. As an adaptive model, ART easily makes this prediction without sacrificing internal consistency: It uses any available information to achieve stable internal representations of signals.

Copious evidence suggests that lexical memory includes both abstract and episodic representations. Consider *perceptual learning* in speech: Because the acoustic realization of any phoneme differs both within and across speakers, listeners must map varied spoken input onto *intended* phonemic categories (Apfelbaum et al., 2014), and they can quickly adjust phonemic categories across speakers (Dahan et al., 2008; Eisner & McQueen, 2005; Kraljic & Samuel, 2005, 2007; Norris, McQueen, & Cutler, 2003). Listeners use abstract, segmental knowledge to map input onto stable categories, and speaker-specific knowledge to guide the mapping.

Although the present study examined repetition priming, a similar convergence of abstract and episodic representations is observed in "higher" linguistic behavior. Altmann and Kamide (1999, 2009; Kamide, Altmann, & Haywood, 2003) used a visual world paradigm to examine anticipatory eye movements in sentence comprehension. In this paradigm, people make anticipatory saccades to depicted objects, almost in real-time with ongoing sentence processing. Recently, Kamide (2012) used this paradigm to test whether the interpretation of ambiguous clauses is speaker-specific. Participants viewed scenes depicting various objects (e.g., a man, a girl, a motorbike, a carousel) while hearing sentences with structurally ambiguous relative clauses (e.g., *"The uncle of the girl who will ride the carousel/motorbike is from France"*). Whereas one speaker consistently spoke sentences with *who* modifying the first noun phrase (such that the uncle rode the motorbike), another

consistently modified the second noun phrase. Upon hearing new sentences produced by those speakers, participants made anticipatory eye movements to whichever image was consistent with each speaker's "attachment style." Voice-specific memory helped resolve syntactic ambiguity in real-time, an example of abstract and specific representations working together in perception.

To accommodate the joint influences abstract and episodic representations on word perception, Goldinger (2007) cited the *complementary learning systems* (CLS) theory (McClelland, McNaughton, & O'Reilly, 1995; Norman & O'Reilly, 2003), which posits reciprocal hippocampal and cortical structures to support fast learning of specific instances and the slow appreciation of generalities. In CLS, the hippocampus is theorized to rapidly encode idiosyncratic events. In contrast, the neocortex receives input from the hippocampus (among other structures) and slowly derives stable prototypes. The system is hybrid; the hippocampal and cortical systems are interdependent. Upon word presentation (for example), the cortical system provides the categorical structure necessary for linguistic processing. Simultaneously, the hippocampus supports episodic encoding, leading to voice-specific priming effects, perceptual learning, and the experience of memory. It is unclear whether such a model will fully explain the joint effects of abstract and episodic representations in speech, but it is abundantly clear that both forms of lexical memory matter. When a person encounters an old friend, the experience is both abstract and specific: All standard features (eyes, chin, etc.) are naturally appreciated as part of general face processing. At the same time, this particular person is an old friend, whose face triggers countless specific memories. From a theoretical perspective, our goal should be to develop accounts that elegantly combine cognitive stability and plasticity. Stability is critical for basic functions (e.g., reading, recognizing objects, etc.), but plasticity is the foundation for our own personal life stories.

Acknowledgments

This research was supported by NICHD Grant R01 HD075800-02, awarded to S.D. Goldinger. We thank Monica Poore, Kyle Brady, Geoff McKinley and Gabrielle Muniz for their assistance with data collection, and Melissa Miola, Tresa Marchi, Joel Chabrier, Rachelle Friedman, Suhani Mehrotra, Jessie Moecia, and Jorin Larsen for assistance creating the stimuli.

References

Abercrombie, D. Elements of general phonetics. Chicago: Aldine Publishing Co; 1967.

Allen JS, Miller JL. Listener sensitivity to individual talker differences in voice-onset-time. Journal of the Acoustical Society of America. 2004; 115:3171–3183. [PubMed: 15237841]

Allopenna PD, Magnuson JS, Tanenhaus MK. Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. Journal of Memory and Language. 1998; 38:419–439.

Altmann GTM, Kamide Y. Incremental interpretation at verbs: restricting the domain of subsequent reference. Cognition. 1999; 73:247–264. [PubMed: 10585516]

Altmann GTM, Kamide Y. Discourse-mediation of the mapping between language and the visual world: Eye movements and mental representation. Cognition. 2009; 111:55–71. [PubMed: 19193366]

Apfelbaum K, Bullock-Rest N, Rhone A, Jongman K, McMurray B. Contingent categorization in speech perception. Language, Cognition and Neuroscience. 2014; 29:1070–1082.

Balota DA, Spieler DH. Word frequency, repetition, and lexicality effects in word recognition tasks: Beyond measures of central tendency. Journal of Experimental Psychology: General. 1999; 128:32–55. [PubMed: 10100390]

Balota DA, Yap MJ, Cortese MJ, Hutchison KA, Kessler B, Loftis B, Treiman R. The English Lexicon Project. Behavior Research Methods. 2007; 39:445–459. [PubMed: 17958156]

Bechtel, W.; Abrahamsen, AA. Connectionism And The Mind : Parallel Processing, Dynamics, And Evolution In Networks. 2nd ed.. Malden, MA: Blackwell; 2002.

Bowers JS. In defense of abstractionist theories of repetition priming and word identification. Psychonomic Bulletin and Review. 2000; 7:83–99. [PubMed: 10780021]

Bradlow AR, Nygaard LC, Pisoni DB. Effects of talker, rate, and amplitude variation on recognition memory for spoken words. Perception & Psychophysics. 1999; 61(2):206–219. [PubMed: 10089756]

Campeanu S, Craik FIM, Alain C. Voice congruency facilitates word recognition. PLoS ONE. 2013; 8(3):e58778. [PubMed: 23527021]

Chomsky, N. The Minimalist Program. MIT Press; 1995.

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(3):521–533.

Craik FIM, Kirsner K. The effect of speaker's voice on word recognition. Quarterly Journal of Experimental Psychology. 1974; 26:274–284.

Creel SC, Aslin RN, Tanenhaus MK. Heeding the voice of experience: The role of talker variation in lexical access. Cognition. 2008; 106:633–664. [PubMed: 17507006]

Creel SC. Preschoolers' flexible use of talker information during word learning. Journal of Memory and Language. 2014; 73:81–98.

Creel SC, Tumlin MA. On-line acoustic and semantic interpretation of talker information. Journal of Memory and Language. 2011; 65:264–285.

Creel SC, Tumlin MA. On-line recognition of music is influenced by relative and absolute pitch information. Cognitive Science. 2012; 36:261–285. [PubMed: 22050775]

Dahan D, Drucker SJ, Scarborough RA. Talker adaptation in speech perception: Adjusting the signal or the representations? Cognition. 2008; 108:710–718. [PubMed: 18653175]

Dahan D, Magnuson JS, Tanenhaus MK. Time course of frequency effects in spoken-word recognition: Evidence from eye movements. Cognitive Psychology. 2001; 42:317–367. [PubMed: 11368527]

Dahan D, Tanenhaus MK. Continuous mapping from sound to meaning in spoken-language comprehension: Immediate effects of verb-based thematic constraints. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2004; 30:498–513.

Eisner F, McQueen JM. The specificity of perceptual learning in speech processing. Perception & Psychophysics. 2005; 67:224–238. [PubMed: 15971687]

Faust ME, Balota DA, Spieler DH, Ferraro FR. Individual differences in information processing rate and amount: Implications for group differences in response latency. Psychological Bulletin. 1999; 125:777–799. [PubMed: 10589302]

Francis G. Too good to be true: Publication bias in two prominent studies from experimental psychology. Psychonomic Bulletin & Review. 2012; 19:151–156. [PubMed: 22351589]

Francis G. Replication, statistical consistency, and publication bias. Journal of Mathematical Psychology. 2013; 57:153–169.

Fujimoto M. The effect of voice variation on the nature of the representation of speech and recognition memory: Evidence from form-based priming. University at Buffalo Working Papers on Language and Perception. 2003; 2:87–163.

Gagnepain P, Chételat G, Landeau B, Dayan J, Eustache F, Lebreton K. Spoken word memory traces within the human auditory cortex revealed by repetition priming and functional magnetic resonance imaging. Journal of Neuroscience. 2008; 2:5281–5289. [PubMed: 18480284]

Geiselman RE, Bellezza FS. Long-term memory for speaker's voice and source location. Memory & Cognition. 1976; 4:483–489. [PubMed: 21286971]

Geiselman RE, Bellezza FS. Incidental retention of speaker's voice. Memory & Cognition. 1977; 5:658–665. [PubMed: 24203282]

Goh WD. Talker variability and recognition memory: Instance-specific and voice-specific effects. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2005; 31:40–53.

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]

Goldinger, SD. Proceedings of 2007 International Congress on Phonetic Sciences. Saarbrucken, Germany: 2007. A complementary-systems approach to abstract and episodic speech perception; p. 49-54.

Goldinger SD, Azuma T. Puzzle-solving science: The quixotic quest for units in speech perception. Journal of Phonetics. 2003; 31:205–320.

Goldinger SD, Azuma T. Episodic memory reflected in printed word naming. Psychonomic Bulletin & Review. 2004; 11(4):716–722. [PubMed: 15581123]

Goldinger SD, Pisoni DB, Logan JS. On the nature of talker variability effects ion serial recall of spoken word lists. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1991; 17:152–162.

González J, McLennan CT. Hemispheric differences in indexical specificity effects in spoken word recognition. Journal of Experimental Psychology: Human Perception and Performance. 2007; 33:410–424. [PubMed: 17469976]

Grossberg S. How does a brain build a cognitive code? Psychological Review. 1980; 87:1–51. [PubMed: 7375607]

Grossberg S. The link between brain learning, attention, and consciousness. Consciousness and Cognition. 1999; 8:1–44. [PubMed: 10072692]

Grossberg S. Resonant neural dynamics of speech perception. Journal of Phonetics. 2003; 31:423–445.

Grossberg S, Boardman I, Cohen C. Neural dynamics of variable-rate speech categorization. Journal of Experimental Psychology: Human Perception and Performance. 1997; 23:418–503.

Grossberg S, Myers CW. The resonant dynamics of speech perception: Interword integration and duration-dependent backward effects. Psychological Review. 2000; 107:735–767. [PubMed: 11089405]

Hinton, GE.; Anderson, JA. Parallel Models Of Associative Memory. Hillsdale, NJ: Lawrence Erlbaum Associates; 1981.

Hintzman DL. "Schema abstraction" in a multiple-trace memory model. Psychological Review. 1986; 93:411–428.

Hintzman DL. Judgments of frequency and recognition memory in a multiple-trace memory model. Psychological Review. 1988; 95:528–551.

Houston D, Jusczyk P. The role of talker-specific information in word segmentation by infants. Journal of Experimental Psychology: Human Perception and Performance. 2000; 26:230–257.

Houston D, Jusczyk P. Infants' long-term memory for the sound patterns of words and voices. Journal of Experimental Psychology: Human Perception and Performance. 2003; 29:1143–1154. [PubMed: 14640835]

Hutchison KA, Balota DA, Cortese MJ, Watson JM. Predicting semantic priming at the item level. Quarterly Journal of Experimental Psychology. 2008; 61:1036–1066.

Jacoby LL, Brooks LR. Nonanalytic cognition: Memory, perception, and concept-learning. Psychology of Learning and Motivation: Advances in Research and Theory. 1984; 18:1–47.

Jacoby LL, Dallas M. On the relationship between autobiographical memory and perceptual-learning. Journal of Experimental Psychology: General. 1981; 110:306–340. [PubMed: 6457080]

Jacoby LL, Hayman CAG. Specific visual transfer in word identification. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1987; 13:456–463.

Johnson K. Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. Journal of Phonetics. 2006; 34:485–499.

Johnsrude I, Mackey A, Hakyemez H, Alexander A, Trang HP, Carlyon RP. Swinging at a cocktail party: Voice familiarity aids speech perception in the presence of a competing voice. Psychological Science. 2013; 24:1995–2004. [PubMed: 23985575]

Kamide Y. Learning individual talkers' structural preferences. Cognition. 2012; 124:66–71. [PubMed: 22498776]

Kamide Y, Altmann GTM, Haywood SL. The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. Journal of Memory and Language. 2003; 49:133–156.

Kirk KI, Pisoni DB, Miyamoto RC. Effects of stimulus variability on speech perception in listeners with hearing impairment. Journal of Speech, Language, and Hearing Research. 1997; 40:1395–1405.

Kraljic T, Samuel AG. Perceptual learning in speech: Is there a return to normal? Cognitive Psychology. 2005; 51:141–178. [PubMed: 16095588]

Kraljic T, Samuel AG. Perceptual adjustments to multiple speakers. Journal of Memory and Language. 2007; 56:1–15.

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

Lahiri A, Marslen-Wilson W. The mental representation of lexical form: A phonological approach to the recognition lexicon. Cognition. 1991; 38:245–294. [PubMed: 2060271]

Lewis MB. Face-space-R: Towards a unified account of face recognition. Visual Cognition. 2004; 11:29–69.

Loftus GR. On interpretation of interactions. Memory & Cognition. 1978; 6:312–319.

Logan GD. Toward an instance theory of automatization. Psychological Review. 1988; 95:492–527.

Luce PA, Lyons EA. Specificity of memory representations for spoken words. Memory & Cognition. 1998; 26:708–715. [PubMed: 9701963]

Luce PA, Pisoni DB. Recognizing spoken words: The neighborhood activation model. Ear and Hearing. 1998; 19:1–36. [PubMed: 9504270]

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]

Magnuson JS, Tanenhaus MK, Aslin RN, Dahan D. The time course of spoken word learning and recognition: Studies with artificial lexicons. Journal of Experimental Psychology: General. 2003; 132:202–227. [PubMed: 12825637]

Maibauer AM, Markis TA, Newell J, McLennan CT. Famous talker effects in spoken word recognition. Attention, Perception, & Psychophysics. 2014; 76:11–18.

Marslen-Wilson W, Warren P. Levels of perceptual representation and process in lexical access: Words, phonemes, and features. Psychological Review. 1994; 101(4):653–675. [PubMed: 7984710]

Martin C, Mullennix JW, Pisoni DB, Summers W. Effects of talker variability on recall of spoken word lists. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1989; 15:676–684.

McClelland JL, McNaughton BL, O'Reilly RC. Why there are complementary learning-systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. Psychological Review. 1995; 102:419–457. [PubMed: 7624455]

McClelland JL, Rumelhart D. Distributed memory and the representation of general and specific information. Journal of Experimental Psychology: General. 1985; 114:159–188. [PubMed: 3159828]

McLennan CT, González J. Examining talker effects in the perception of native- and foreign-accented speech. Attention, Perception, and 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(2): 306–321.

McLennan, CT. Proceedings of the 16th International Congress of Phonetic Sciences. Saarbrücken, Germany: 2007. Challenges facing a complementary-systems approach to abstract and episodic speech perception; p. 67-70.

McMurray B, Jongman A. What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations. Psychological Review. 2011; 118:219–246. [PubMed: 21417542]

McQueen JM, Norris D, Cutler A. Are there really interactive processes in speech perception? Trends in Cognitive Sciences. 2006; 10:533–533. [PubMed: 17067845]

Meehan EF, Pilotti M. Auditory priming in an implicit memory task that emphasizes surface processing. Psychonomic Bulletin & Review. 1996; 3:495–498. [PubMed: 24213983]

Mitterer H, Ernestus M. The link between speech perception and production is phonological and abstract: Evidence from the shadowing task. Cognition. 2008; 109:168–173. [PubMed: 18805522]

Mullennix JW, Pisoni DB. Stimulus variability and processing dependencies in speech perception. Perception & Psychophysics. 1990; 47:379–390. [PubMed: 2345691]

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]

Namy L, Nygaard LC, Sauerteig D. Gender differences in vocal accommodation: The role of perception. Journal of Language and Social Psychology. 2002; 21:422–432.

Nearey TM. Speech perception as pattern recognition. The Journal of the Acoustical Society of America. 1997; 101:3241–3254. [PubMed: 9193041]

Newman R, Evers S. The effect of talker familiarity on stream segregation. Journal of Phonetics. 2007; 35:85–103.

Nielsen K. Specificity and abstractness of VOT imitation. Journal of Phonetics. 2011; 39:132–142.

Norman KA, O'Reilly RC. Modeling hippocampal and neocortical contributions to recognition memory: A complementary-learning-systems approach. Psychological Review. 2003; 110:611–646. [PubMed: 14599236]

Norris D. Shortlist: A connectionist model of continuous speech recognition. Cognition. 1994; 52:189–234.

Norris D, McQueen JM, Cutler A. Perceptual learning in speech. Cognitive Psychology. 2003; 47:204–238. [PubMed: 12948518]

Nusbaum, HC.; Morin, TM. Paying attention to differences among talkers. In: Tohkura, Y.; Sagisaka, Y.; Vatikiotis-Bateson, E., editors. Speech perception, speech production, and lingustic structure. Tokyo: OHM; 1992. p. 113-134.

Nygaard LC, Pisoni DB. Talker-specific learning in speech perception. Perception & Psychophysics. 1998; 60:355–376. [PubMed: 9599989]

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

Orfanidou E, Davis MH, Ford MA, Marslen- Wilson WD. Perceptual and response com- ponents in repetition priming of spoken words and pseudowords. The Quarterly Journal of Experimental Psychology. 2011; 64:96–121. [PubMed: 20509097]

Pallier C, Colome A, Sebastian-Galles N. The influence of native-language phonology on lexical access: Exemplar-based versus abstract lexical entries. Psychological Science. 2001; 12:445–449. [PubMed: 11760129]

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(2):309–328.

Pardo J. On phonetic convergence during conversational interaction. Journal of the Acoustical Society of America. 2006; 119:2382–2393. [PubMed: 16642851]

Pierrehumbert, J. Exemplar dynamics: Word frequency, lenition, and contrast. In: Bybee, J.; Hopper, P., editors. Frequency effects and the emergence of lexical structure. John Benjamins, Amsterdam: 2001. p. 137-157.

Pilotti M, Bergman ET, Gallo DA, Sommers M, Roediger HL. Direct comparison of auditory implicit memory tests. Psychonomic Bulletin & Review. 2000; 7:347–353. [PubMed: 10909144]

Pilotti M, Beyer T. Perceptual and lexical components of auditory repetition priming in young and older adults. Memory & Cognition. 2002; 30:226–236. [PubMed: 12035884]

Pilotti M, Meade ML, Gallo DA. Implicit and explicit measures of memory for perceptual information in young adults, healthy older adults, and patients with Alzheimer's Disease. Experimental Aging Research. 2003; 29:15–32. [PubMed: 12735079]

Pisoni DB. Long-term memory in speech perception: Some new findings on talker variability, speaking rate, and perceptual learning. Speech Communication. 1993; 13:109–125. [PubMed: 21461185]

Ratcliff R, Rouder JN. Modeling response times for two-choice decisions. Psychological Science. 1998; 9:347–356.

Rosenblum LD, Miller RM, Sanchez K. Lipread me now, hear me better later: Crossmodal transfer of talker familiarity effects. Psychological Science. 2007; 18:392–396. [PubMed: 17576277]

Ryalls BO, Pisoni DB. The effect of talker variability on word recognition in preschool children. Developmental Psychology. 1997; 33:441–452. [PubMed: 9149923]

Scarborough DL, Cortese C, Scarborough HS. Frequency and repetition effects in lexical memory. Journal of Experimental Psychology: Human Perception and Performance. 1977; 3:1–17.

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.

Sheffert S. Voice-specificity effects on auditory word priming. Memory & Cognition. 1998a; 26:591–598. [PubMed: 9610127]

Sheffert S. Contributions of surface and conceptual information to recognition memory. Perception & Psychophysics. 1998b; 60:1141–1152. [PubMed: 9821776]

Sheffert S, Fowler C. The effects of voice and visible speaker change on memory for spoken words. Journal of Memory and Language. 1995; 34:665–685.

Shockley K, Sabadini L, Fowler C. Imitation in shadowing words. Perception & Psychophysics. 2004; 66:422–429. [PubMed: 15283067]

Smith R, Hawkins S. Production and perception of speaker-specific phonetic detail at word boundaries. Journal of Phonetics. 2012; 40:213–233.

Smits R. Hierarchical categorization of coarticulated phonemes: A theoretical analysis. Perception & Psychophysics. 2001; 63:1109–1139. [PubMed: 11766939]

Sommers MS. The structural organization of the mental lexicon and its contribution to age-related declines in spoken-word recognition. Psychology and Aging. 1996; 11:333–341. [PubMed: 8795062]

Stevens KN. Toward a model for lexical access based on acoustic landmarks and distinctive features. Journal of the Acoustical Society of America. 2002; 111:1872–1891. [PubMed: 12002871]

Sumner M. The role of variation in the perception of accented speech. Cognition. 2011; 119:131–136. [PubMed: 21144500]

Sumner M, Samuel AG. The effect of experience on the perception and representation of dialect variants. Journal of Memory and Language. 2009; 60:487–501.

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]

Underwood BJ. Attributes of memory. Psychological Review. 1969; 76:559–573.

Valentine T. A unified account of the effects of distinctiveness, inversion and race in face recognition. Quarterly Journal of Experimental Psychology. 1991; 43A:161–204. [PubMed: 1866456]

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]

Wade T, Dogil G, Shütze H, Walsh M, Möbius B. Syllable frequency effects in a context-sensitive segment production model. Journal of Phonetics. 2010; 38:227–239.

Wagenmakers E-J, Brown S. On the linear relation between the mean and the standard deviation of a response time distribution. Psychological Review. 2007; 114:830–841. [PubMed: 17638508]

Wagenmakers E-J, Krypotos A-M, Criss AH, Iverson G. On the interpretation of removable interactions: A survey of the field 33 years after Loftus. Memory & Cognition. 2012; 40:145–160. [PubMed: 22069144]

Yonan CA, Sommers MS. The effects of talker familiarity on spoken word identification in younger and older adults. Psychology and Aging. 2000; 15:88–99. [PubMed: 10755292]

Appendix A

Results for “standard” lexical decision and semantic classification, using keyboard input

Description

In order to validate that our novel eye-tracking procedure faithfully approximates standard lexical decision and semantic classification tasks, we conducted two small-scale experiments. Both were identical to identical to Experiment 1, but used keyboard input (the ‘f’ and ‘j’ keys), rather than the find-and-click methods of the main experiments. There were 16 participants (mean age 19.4 years, 11 women) for lexical decision, and 21 participants (mean age 19.3 years, 12 women) for semantic classification.

Lexical Decision

The mean RTs for the second block of trials, along with calculated priming and voice effects are shown in Table A1. The results were analyzed in a 3×3 subject-based, RM ANOVA, with factors Stimulus Type (HF, LF, NW) and Repetition Type (New, SV, DV), with planned comparisons to evaluate priming and voice effects. There were reliable main effects of both Stimulus Type, $F(2, 14) = 8.1$, $p < .01$, $\eta^2_p = .19$, and Repetition Type, $F(2, 14) = 5.9$, $p < .05$, $\eta^2_p = .11$, both in their expected directions. HF words (801 ms) were classified faster than LF words (821 ms), which were classified faster than Nonwords (884 ms). Although there were reliable simple effects for SV words, relative to new words, few planned comparisons were reliable. Voice effects were marginal for both HF words (23 ms, $p = .08$) and LF words (20 ms, $p = .11$), and reliable for Nonwords (29 ms, $p = .04$).

With respect to the question of interest, data collected with our novel eye-tracking method appear generally similar to data collected using the standard, keyboard method. TTI results (see Table 2) were 164 ms faster than keyboard RTs, but with comparable frequency effects (39 vs. 21 ms, respectively). Both priming effects and voice effects were stronger in the TTI data, relative to the keyboard data (by 47 and 33 ms, respectively). TTF results were 124 ms slower than keyboard RTs, again with comparable frequency effects (41 vs. 21 ms, respectively). Both priming and voice effects were slightly larger in the TTF results than the keyboard results (by 23 and 19 ms, respectively). Given the different sample sizes and response modes, we do not consider these differences further: The main point is that our novel procedure produces lexical decision results that are broadly similar to the more standard paradigm.

Table A1

RTs (in ms) to new and repeated items (with resultant priming measures) in lexical decision, as a function of stimulus type and voice. “Priming” denotes differences, relative to new words, and “Voice Effects” compare SV versus DV priming.

	HF Words	Priming	LF Words	Priming	NW	Priming
New:	819		843		890	
SV:	780	39*	801	42*	866	23, ns
DV:	803	16, ns	819	22, ns	895	−5, ns
Voice Effect:		23, ns		20, ns		29*

* $p < .05$

Semantic Classification

The mean RTs for the second block of trials, along with calculated priming and voice effects are shown in Table A2. The results were analyzed in a 2×3 RM ANOVA, with factors Frequency (HF, LF) and Repetition Type (New, SV, DV), with planned comparisons to evaluate priming and voice effects. There were main effects of Frequency, $F(1, 20) = 20.7$, $p < .001$, $\eta^2_p = .34$, and Repetition Type, $F(2, 19) = 11.2$, $p < .01$, $\eta^2_p = .18$, both in their expected directions. HF words (824 ms) were classified faster than LF words (861 ms). Priming was robust for both SV and DV words. The voice effects were marginal for HF words (19 ms, $p = .10$) and reliable for LF words (30 ms, $p = .02$).

As with the lexical decision data, the results were broadly consistent with those from the eye-tracking procedure, but with stronger priming and voice effects in the eye-tracking results. TTI results (see Table 3) were 110 ms faster than keyboard RTs, with slightly stronger priming (by 32 ms) and voice effects (by 15 ms). TTF results were 192 ms slower than keyboard RTs, but both priming and voice effects were again stronger, relative to the keyboard RTs (by 44 and 37 ms, respectively).

Table A2

RTs (in ms) to new and repeated items (with resultant priming measures) in semantic classification, as a function of stimulus type and voice. “Priming” denotes differences, relative to new words, and “Voice Effects” compare SV versus DV priming.

	HF Words	Priming	LF Words	Priming
New:	860		905	
SV:	796	64*	820	85**
DV:	815	45*	859	46*
Voice Effect:		19, ns		39*

* $p < .05$,
** $p < .01$

Appendix B

Separate results for each direction of priming in Experiment 2

In the main text, the results of Experiment 2 are shown in Table 4, which shows mean RTs for all participants' second block of trials, collapsed across task order. To allow full appreciation for the results, Table B1 shows results for participants who performed SC followed by LD, and Table B2 shows results for participants who performed LD followed by SC.

Table B1

Lexical decision RTs to new and repeated items (with resultant priming measures) as a function of stimulus type and voice, Experiment 2. Task order was SC followed by LD.

	HF Words	*Priming*	*Voice Effect*	LF Words	*Priming*	*Voice Effect*
Mouse-Click RTs						
New Item	1635 (59)			1660 (48)		
Same Voice	1516 (52)	119***		1529 (50)	131***	
Different Voice	1536 (52)	99***	20, ns	1602 (40)	58*	73*
Time-to-Fixate						
New Item	901 (48)			939 (35)		
Same Voice	808 (39)	93**		840 (40)	99***	
Different Voice	840 (32)	61*	32*	891 (30)	48**	51*
Time-to-Initiate						
New Item	669 (33)			720 (48)		
Same Voice	600 (38)	69***		637 (41)	83**	
Different Voice	635 (22)	34*	35*	671 (35)	49*	34*

Notes: Standard errors shown in parentheses. "Priming" denotes the contrast of repeated items versus new words. "Voice Effect" denotes the comparison of same- and diffferent-voice repetitions.

* $p < .05$;

** $p < .01$;

*** $p < .001$

For the LD data, results for the separate dependent measures (mouse RTs, TTF, and TTI) were analyzed in 2×3 RM ANOVAs (based on participants), with factors Frequency (HF, LF) and Repetition Type (New, SV, DV), and planned comparisons to evaluate priming and voice effects. Nonword results were not analyzed, as they could not receive priming from the preceding SC block. We report only key effects of interest here, avoiding full treatment of all interactions. There were reliable main effects of Frequency in the TTI and TTF measures (both $F(1, 23) > 10.5$, $p < .01$), but it was not reliable for the mouse RTs. Main effects of Repetition Type were reliable for all three measures (all $F(1, 23) > 12.0$, $p < .01$). As shown in Table B1, priming effects were reliable in every observation, and voice effects were reliable in every case except mouse RTs for HF words.

Table B2

Semantic classification RTs to new and repeated items (with resultant priming measures) as a functionof stimulus type and voice, Experiment 2. Task order was LD followed by SC.

	HF Words	*Priming*	*Voice Effect*	LF Words	*Priming*	*Voice Effect*
Mouse-Click RTs						
New Item	2094 (78)			2036 (77)		
Same Voice	2033 (70)	57*		1922 (66)	114**	
Different Voice	2041 (67)	49*	8, ns	1990 (61)	46*	68**
Time-to-Fixate						
New Item	1151 (51)			1060 (44)		
Same Voice	1064 (46)	87**		1017 (38)	43*	
Different Voice	1099 (52)	52*	35, ns	1044 (36)	16, ns	27, ns
Time-to-Initiate						
New Item	903 (41)			977 (42)		
Same Voice	870 (35)	33*		909 (36)	68**	
Different Voice	925 (39)	−22, ns	55**	952 (40)	25, ns	43*

Notes: Standard errors shown in parentheses. "Priming" denotes the contrast of repeated items versus new words. "Voice Effect" denotes the comparison of same- and diffferent-voice repetitions.

* $p < .05$;

** $p < .01$;

*** $p < .001$

For the SC data, results for the separate dependent measures were also analyzed in 2×3 RM ANOVAs (based on participants), with factors Frequency (HF, LF) and Repetition Type (New, SV, DV), with planned comparisons to evaluate priming and voice effects. There was a reliable main effect of Frequency in the Mouse RTs, $F(1, 23) = 15.5$, $p < .01$, $\eta^2_p = .17$, but it was “backwards,” with faster RTs to low-frequency words. The Frequency effect was marginal, in the usual direction, for the TTF results, $F(1, 23) = 2.9$, $p = .07$, and it was reliable for the TTI results, $F(1, 23) = 19.0$, $p < .001$, $\eta^2_p = .31$. Main effects of Repetition Type were reliable for several measures (LF and HF mouse RTs, HF TTF, and LF TTI measures; all $F(1, 23) > 6.5$, $p < .05$). As shown in Table B2, priming effects were generally robust (with several exceptions). Voice effects were only reliable in three cases (see Table B2), mainly in the TTI results.

In examining the results separately, the key question of interest is whether voice effects were strongly modulated by task order. To assess this, we conducted combined, mixed-model 2×2×3 ANOVAs (one for each dependent measure). Task Order (LD-SC, SC-LD) was the between-subjects factor, with Frequency (HF, LF) and Repetition Type (New, SV, DV) as within-subject factors. We focus only on novel findings related to Task Order. For all three dependent measures (mouse RTs, TTF, and TTI), there were large main effects of Task Order (all $F(2, 46) > 59.0$, $p < .001$), reflecting slower performance in SC, relative to LD. For both Mouse RTs and TTF, there were reliable interactions of Task Order × Frequency (both $F(2, 46) > 11.5$, $p < .01$), reflecting “backward” frequency effects in the LD-SC task order. No such pattern was observed for TTI. No reliable interactions were observed

between Task Order × Repetition Type (all $F < 1.0$). Similarly, no reliable interactions were observed for Task Order × Frequency X Repetition Type (again, all $F < 1.0$). Overall, there was no evidence that Task Order affected priming or voice effects.

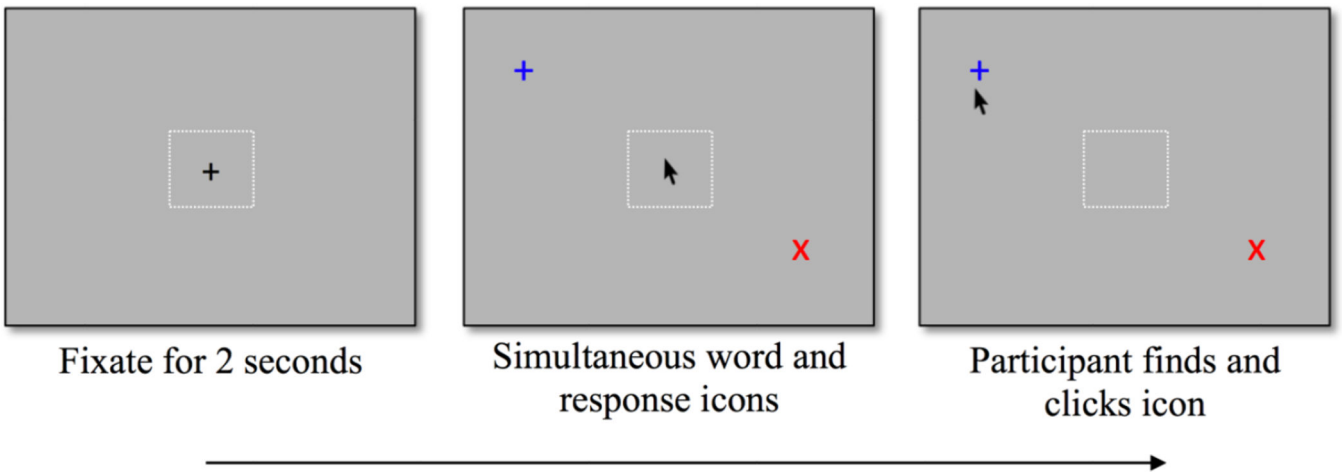

Figure 1.

Schematic trial outline. Each trial began with a gaze-contingent 2000-ms fixation cross, followed by the onset of a spoken word for either lexical decision (LD) or semantic classification (SC). 'Word' ('larger than a toaster') decisions were made by locating and clicking a blue '+'; 'nonword' ('smaller than a toaster') decisions were made by clicking a red '×'. Response options were randomly located within the same visual half-field throughout the experiment, but changed locations in every trial. The dashed box in the center did not appear in the procedure, but is shown to illustrate a buffer zone where response icons could not appear.

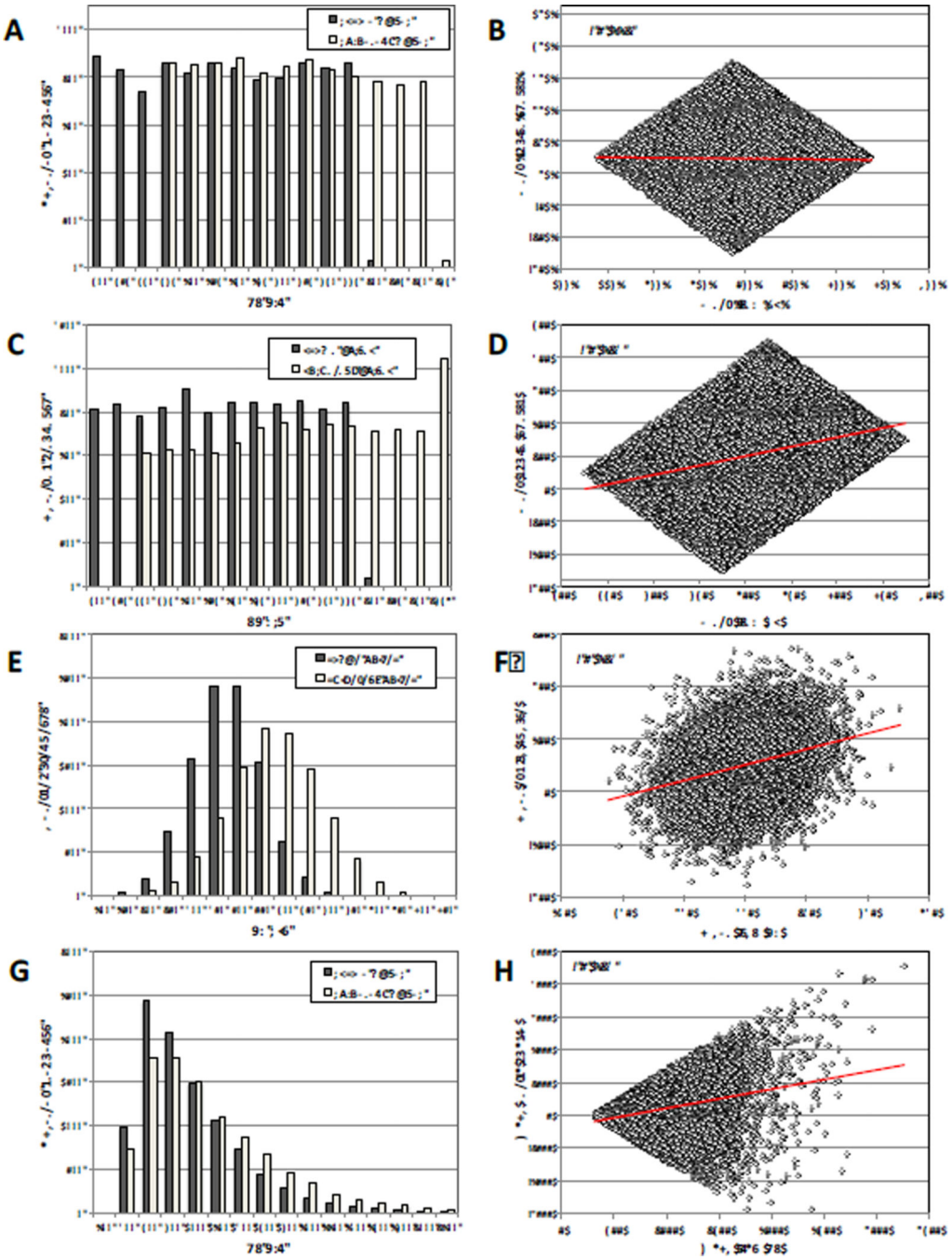

Figure 2.

Simulated sampling of “same-voice” and “different-voice” RTs, showing the relationship that emerges between item RT and voice effects. Panels A, C, E, and G show RT frequency distributions for hypothetical SV and DV words, sampled from flat distributions with equal variance, flat distributions with unequal variance, Gaussian distributions with unequal variance, and Weibull distributions with unequal variance (respectively). Panels B, D, F, and H show the associations that emerge between mean item RTs (SV+DV/2) and the size of voice effects (DV-SV).

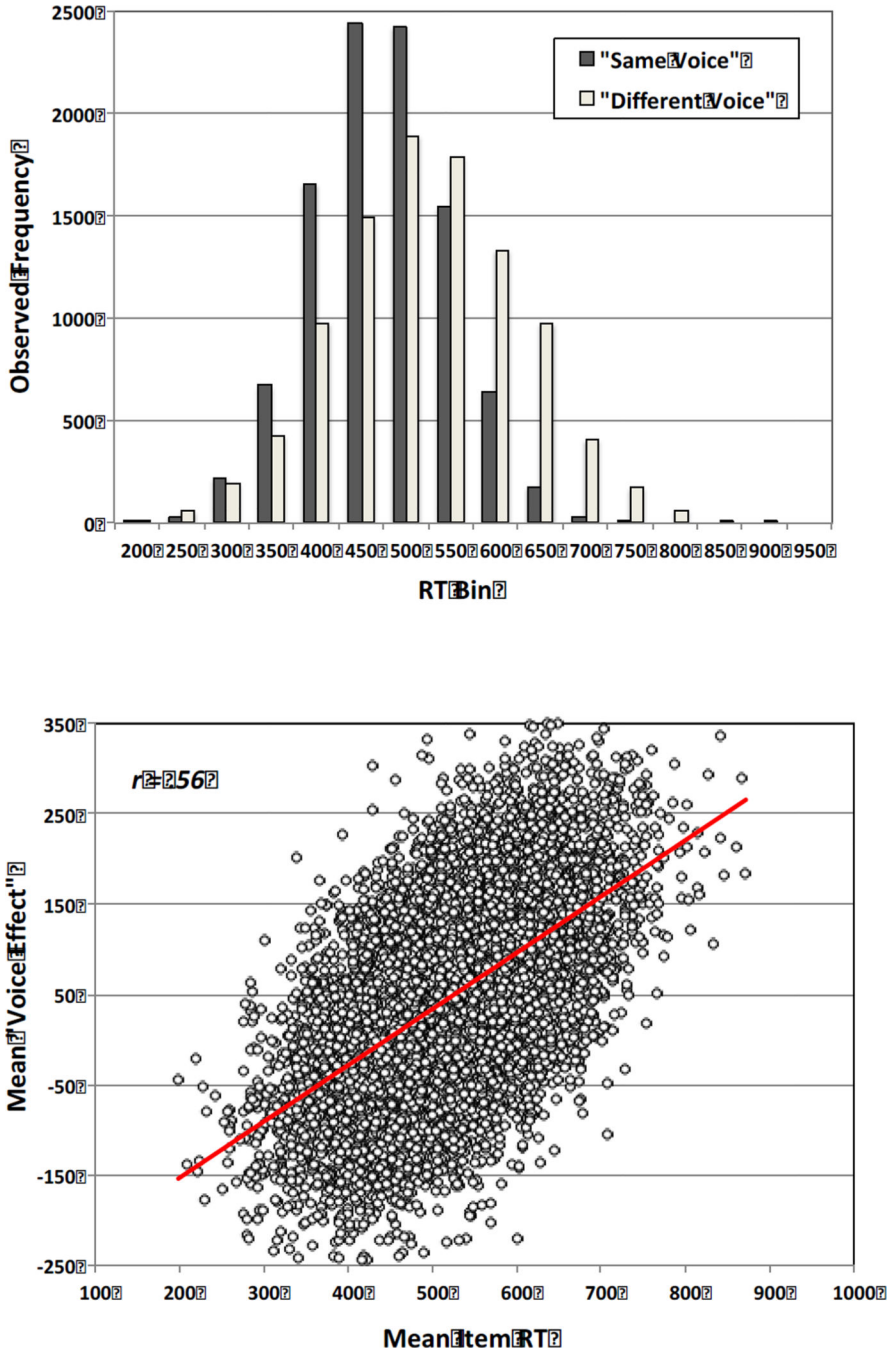

Figure 3.

Simulated sampling of “same-voice” and “different-voice” RTs, with the extra assumption that both versions of each word are loosely related to each other. The upper panel shows RT frequency distributions for hypothetical SV and DV words: SV words were sampled from a Gaussian distribution, and DV versions were created by sampling normally-distributed changes to those base RTs. The lower panel shows the association that emerges between mean item RTs (SV+DV/2) and the size of voice effects (DV-SV).

Table 1

Summary Statistics for the Stimulus Items

		Means				
Item Type	**n**	**Letters**	**Syllables**	**Length (ms)**	**KF†**	**Subtitle Freq.‡**
Exp 1: LD						
HF	40	4.28	1.20	676	414.35	518.86
LF	40	4.82	1.20	717	7.57	11.23
NW	80	5.33	1.52	772	n/a	n/a
Exp 1: SC						
HF	70	4.47	1.21	681	87.57	125.09
LF	70	4.78	1.32	693	10.35	9.81
Exp 2						
HF	56	4.48	1.20	680	84.10	115.21
LF	56	4.91	1.35	697	11.35	10.42
NW	80	5.33	1.52	772	n/a	n/a

All data were obtained from Balota et al. (2007).

†Kuçera & Francis (1967)

‡Brysbaert & New (2009)

Table 2

RTs to new and repeated items (with resultant priming measures) in LDT, as a function of stimulus type and voice, Experiment 1.

	HF Words	*Priming*	*Voice Effect*	LF Words	*Priming*	*Voice Effect*	Nonwords	*Priming*	*Voice Effect*
Mouse-Click RTs									
New Item	1621 (31)			1763 (40)			1848 (33)		
Same Voice	1565 (28)	56**		1626 (31)	137***		1785 (29)	63**	
Different Voice	1605 (44)	17, ns	40, ns	1678 (68)	85**	52, ns	1804 (37)	44*	19, ns
Time-to-Fixate									
New Item	893 (26)			958 (35)			1121 (31)		
Same Voice	845 (21)	48*		870 (21)	88**		1051 (22)	70***	
Different Voice	877 (30)	17, ns	41*	911 (30)	48*	41**	1105 (33)	17, ns	53**
Time-to-Initiate									
New Item	662 (23)			724 (25)			768 (31)		
Same Voice	591 (23)	71***		610 (21)	114***		657 (29)	111***	
Different Voice	637 (23)	25, ns	46***	672 (26)	52***	62***	717 (28)	51***	60***

Notes: Standard errors shown in parentheses. "Priming" denotes the contrast of repeated items versus new items; "Voice Effect" denotes the comparison of same- and diffferent-voice repetitions.

* $p < .05$;

** $p < .01$;

*** $p < .001$

Table 3

RTs to new and repeated items (with resultant priming measures) in Semantic Classification, as a function of stimulus type and voice, Experiment 1.

	HF Words	*Priming*	*Voice Effect*	LF Words	*Priming*	*Voice Effect*
Mouse-Click RTs						
New Item	1803 (47)			1721 (43)		
Same Voice	1586 (35)	217***		1593 (35)	128***	
Different Voice	1625 (36)	178***	39*	1629 (38)	92***	36*
Time-to-Fixate						
New Item	1083 (31)			1136 (39)		
Same Voice	956 (22)	127***		974 (31)	162***	
Different Voice	1014 (22)	69**	58**	1044 (28)	92**	70**
Time-to-Initiate						
New Item	813 (27)			837 (27)		
Same Voice	667 (25)	96***		662 (25)	125***	
Different Voice	704 (25)	59***	37**	701 (26)	86***	39**

Notes: Standard errors shown in parentheses. "Priming" denotes the contrast of repeated items versus new items; "Voice Effect" denotes the comparison of same- and diffferent-voice repetitions.

* $p < .05$;

** $p < .01$;

*** $p < .001$

Table 4

RTs to new and repeated items (with resultant priming measures) as a function of stimulus type and voice, Experiment 2.

	HF Words	*Priming*	*Voice Effect*	LF Words	*Priming*	*Voice Effect*
Mouse-Click RTs						
New Item	1865 (48)			1843 (37)		
Same Voice	1775 (46)	90***		1716 (49)	127***	
Different Voice	1786 (50)	79***	11, ns	1790 (40)	53*	74*
Time-to-Fixate						
New Item	1026 (39)			999 (24)		
Same Voice	936 (30)	90**		928 (29)	71**	
Different Voice	969 (32)	57*	33, ns	972 (30)	27, ns	44*
Time-to-Initiate						
New Item	786 (22)			850 (24)		
Same Voice	734 (20)	52**		773 (26)	77**	
Different Voice	780 (22)	6, ns	46*	811 (23)	39*	38*

Notes: Results are collapsed over the counterbalanced order of tasks. Standard errors shown in parentheses. "Priming" denotes the contrast of repeated items versus new words. "Voice Effect" denotes the comparison of same- and diffferent-voice repetitions.

* $p < .05$;

** $p < .01$;

*** $p < .001$