Available online at www.sciencedirect.com

ScienceDirect

Cognition 106 (2008) 633–664

COGNITION

www.elsevier.com/locate/COGNIT

Heeding the voice of experience: The role of talker variation in lexical access ☆

Sarah C. Creel [a,*], Richard N. Aslin [b], Michael K. Tanenhaus [b]

[a] *Department of Psychology, University of Pennsylvania, 3401 Walnut St. Room 302C, Philadelphia, PA 19104, USA*

[b] *University of Rochester, Rochester, USA*

Received 15 June 2006; revised 20 March 2007; accepted 25 March 2007

Abstract

Two experiments used the head-mounted eye-tracking methodology to examine the time course of lexical activation in the face of a non-phonemic cue, talker variation. We found that lexical competition was attenuated by consistent talker differences between words that would otherwise be lexical competitors. In Experiment 1, some English cohort word-pairs were consistently spoken by a single talker (male *couch*, male *cows*), while other word-pairs were spoken by different talkers (male *sheep*, female *sheet*). After repeated instances of talker-word pairings, words from different-talker pairs showed smaller proportions of competitor fixations than words from same-talker pairs. In Experiment 2, participants learned to identify black-and-white shapes from novel labels spoken by one of two talkers. All of the 16 novel labels were VCVCV word-forms atypical of, but not phonologically illegal in, English. Again, a word was consistently spoken by one talker, and its cohort or rhyme competitor was consistently spoken either by that same talker (same-talker competitor) or the other talker (different-talker competitor). Targets with different-talker cohorts received greater fixation proportions than targets with same-talker cohorts, while the reverse was true for fixations to cohort competitors; there were fewer erroneous selections of competitor referents for different-talker competitors than same-talker competitors. Overall, these results support a view of the lexicon in which entries contain extra-phonemic information. Extensions of the artificial lexicon paradigm and developmental implications are discussed.

☆ This manuscript was accepted under the editorship of Jacques Mehler.

* Corresponding author. Tel.: +1 215 898 4165.
E-mail addresses: creel@psych.upenn.edu, creel@sas.upenn.edu (S.C. Creel).

doi:10.1016/j.cognition.2007.03.013

Keywords: Talker variation; Word recognition; Lexical access; Eye-tracking; Artificial lexicon; Cohort

1. Introduction

The traditional view of speech perception emphasizes the role of phonetic and phonemic categories – stressing the utility of only those acoustic–phonetic cues that conform to the phonological system of a particular natural language – while invoking normalization processes to eliminate non-phonemic (e.g., sub-phonetic, allophonic, or indexical) variation. Different talkers exhibit different fundamental frequencies (Van Lancker, Kreiman, & Wickens, 1985), speaking rates (Van Lancker et al., 1985), voice onset times (Allen, Miller, & DeSteno, 2002), frication noise (Newman, Clouse, & Burnham, 2001), and realizations of vowels (Bradlow, Torretta, & Pisoni, 1996; Clopper & Pisoni, 2004), among myriad other dialectal variations. Such differences occur even after normalizing for speech rate (Allen et al., 2002), and can affect listeners' abilities to apprehend a talker's productions (Newman et al., 2001). However, none of these factors influences word meaning. Phonemic categories, then, are key to identifying lexical entries, and other attributes of the speech signal are viewed as irrelevant and even detrimental to the process of spoken word recognition. This traditional perspective deals effectively with the first of two difficult problems in spoken word recognition: limited storage capacity. By storing only phonemic information, a large number of lexical entries can be represented with a small, compact set of symbols.

A second problem in spoken word recognition is the induction of word categories from disparate acoustic exemplars. A mother saying "dog" produces a token that is acoustically dissimilar to a father saying "dog", or even the mother saying "dog" in different sentential contexts. The novice language learner must become tacitly aware that the acoustic–phonetic variation distinguishing a mother speaking from a father speaking does not cue a lexical difference, while segmental differences (/dag/ vs. /kæt/) described at the level of phonemes do cue lexical identity. Although talker differences can be quite striking, and even though changes in voice quality can indicate global changes in meaning (e.g., sarcasm), they are not known in any natural language to be lexically contrastive (e.g., producing the same segmental information in normal and falsetto voices). Lexical storage that only accepted *lexically contrastive information* that was present in linguistic input would not therefore contain talker information.

However, unless the learner knows *a priori* which differences are relevant to lexical identity and which are non-lexical variations, the mapping of these cues onto the lexicon must be discovered by listening experience. Such a discovery procedure is a formidable problem for the naïve learner. Recent work by Houston and Jusczyk (2000, 2003) suggests that, for early learners, induction is indeed difficult. They found that at 7.5 months of age infants have difficulty recognizing a word they have learned in one voice when that same word is spoken in another voice, whereas 3 months later they readily make this generalization. Barker and Newman (2004) found that 7.5-month-olds were able to learn words from familiar voices (their mothers), but not from unfamiliar voices, when these voices were embedded in distractor speech. These developmental findings are corroborated by a literature in second-language learning that deals with acoustic variability in L2 exposure (Bradlow, Pisoni, Yamada, & Tohkura, 1997; Lively, Logan, & Pisoni, 1993; Logan, Lively, & Pisoni, 1991): it seems that hearing a new phonemic contrast (e.g., /r/-/l/ for native speakers of Japanese) in only one word position (Strange & Dittmann, 1984) or spoken by only one talker (Lively et al., 1993; Logan et al., 1991) does not provide a good basis for perceiving this contrast in other word positions or in the speech of other talkers. Together, these literatures suggest that speech category induction is strongly dependent on the acoustics of the input.

In contrast to models that discard acoustic variation, exemplar-based models of the lexicon (e.g., Goldinger, 1998) predict that lexical representations preserve acoustic detail, even detail that is not used in forming phonetic contrasts. Goldinger (1998) provided evidence from a shadowing task that was consistent with exemplar-specific encoding of words. First, listeners heard numerous words presented only in particular voices. Then, they shadowed word lists including those words. More repetitions of a word in a particular voice during exposure led to faster shadowing times and to greater perceived similarity (judged by new listeners) between the imitation and the token imitated. In other experiments, listeners were familiarized with non-words in a pre-training session using an additional talker who was not employed in the shadowing session. They then heard talker-specific presentations and completed a shadowing session as in the English-word experiment. Again, when participants had heard fewer exposures at pre-training and limited variability of non-words in the training session, there were greater effects of talker-specific presentations (in terms of reaction time and imitation similarity) during the shadowing session.

Additional support for an exemplar theory of lexical representation comes from evidence that sub-phonemically detailed, rather than abstractly phonemic, representations characterize the lexicon. Priming (e.g., Andruski, Blumstein, & Burton, 1994) and eye-tracking studies (e.g., McMurray, Tanenhaus, & Aslin, 2002) have demonstrated gradient voice onset time (VOT) effects on lexical activation. That is, lexical items are not activated in an all-or-none fashion, but are modulated by the degree of match to a (prototypical) VOT. Results like these are consistent with the role of a distributional learning mechanism (Maye, Werker, & Gerken, 2002) that stores acoustic detail: more frequent VOT exemplars are also more prototypical VOT values.

Despite the foregoing evidence in support of sub-phonemic, acoustically detailed storage of words in the lexicon, little work has examined the nature and plasticity of the lexicon to determine what sorts of variability can be considered lexically relevant, given the appropriate distributional properties of the input. Perhaps only those properties of speech that are lexically contrastive in natural languages can be incorporated into the lexicon. Alternatively, if the lexicon is truly an acoustic store, in principle *any* sort of acoustic variability could signify lexical variation. One way to test this hypothesis is to take a non-phonemic property of the speech signal and make it lexically contrastive. If this sort of variation can cue lexical disambiguation, then it would provide strong support for an exemplar-based model of the lexicon (Goldinger, 1998; Hintzman, 1988).

Two recent studies by Kraljic and Samuel (2005, 2006) on perceptual adjustment to a talker's idiosyncratic phoneme productions provide evidence for talker effects on phonemic judgments. Building on an important study by Norris, McQueen, and Cutler (2003), Kraljic and Samuel (2005) exposed listeners to words from a talker who produced an /ʃ/ that was consistently shifted toward /s/ (and vice versa for another set of listeners). This exposure shifted listeners' s/ʃ phoneme boundaries in a categorization task in the direction of the other phoneme, suggesting perceptual adjustment to this talker. This effect was attenuated only by hearing the same talker produce normal s/ʃ, not by the passage of time, by presenting speech in the interim that did not contain /s/ or /ʃ/, or by presenting speech from a different talker that *did* contain /s/ and /ʃ/. Additionally, this perceptual shift did not necessarily transfer to another talker. These results together with related findings by Eisner and McQueen (2006a, 2006b) suggest that the s/ʃ representations involved were robust and talker-specific. Kraljic and Samuel (2006) found a different pattern of results for stop consonants: shifted /d/ or /t/ generalized not only across talker, but also across place of articulation (to b/p). They argue that less "abstract" phoneme categories like s/ʃ may be more likely to vary by talker and to have talker-specific representations than more abstract phoneme categories (d/t). Another interpretation was that the manipulated cue, VOT, had temporal rather than spectral properties as in the s/ʃ study, and temporal properties may be less variable across talkers and places of articulation than spectral ones (see Mirman, McClelland, & Holt, 2006, for an acoustic account of these patterns of generalization). For present purposes, it is sufficient to note that, as in the case of r/l learning for native speakers of Japanese, phonemic information from one talker does not necessarily generalize to another.

In sum, we know that talker variation is perceived, that it can affect task performance, and that it influences learning of novel (or altered) phonemes. However, McLennan and Luce (2005, Experiments 2 & 3) have recently provided evidence that talker effects are slow relative to (and therefore distinct from) lexical effects – that is, talker effects show up only when word-level processing (lexical decision) is slowed down by making the task itself more difficult. Listeners performed a lexical decision task in two blocks of stimuli (Experiment 2). In the first block, some words were presented for the first time, and in the second block, those "primed" words were presented again, either with the *same talker* as in the first block, or a *different talker*. The crucial manipulation was that the lexical decision task was made "easy" or "difficult" by varying the phonotactic probability of the non-words. Non-words are of course less probable than real words, and the more improbable the non-words, the more distinguishable they are from real words. That is, the less word-like the non-words, the easier it is to determine that probable items are words. In the easy condition, words (e.g., *bear*) were much higher in phonotactic probability (in terms of positional segment frequency and biphone frequency) than non-words (e.g., /dʒaʊm/). Contrastingly, in the difficult condition, words (e.g., *book*) and non-words (e.g., /bʊp/) were similar in phonotactic probability, with non-words differing from real words only at the final segment. In both conditions, they found priming (faster reaction times) for previously presented words; importantly, though, they only found a talker effect – i.e., heightened priming for same-talker repetitions – in the difficult condition, where reaction times were slowest (unprimed RTs of 837 ms vs. 800 ms in the easy condition). Because talker effects surfaced only when processing time was extended, McLennan and Luce suggested that talker information has relatively late effects on lexical identification.

An alternate interpretation of McLennan and Luce's (2005) finding of long-term priming for same-talker items is that listeners in the two conditions adopted different strategies in order to perform the task as quickly as possible. In the easy condition, listeners may have learned that they could differentiate words from non-words rapidly based on the frequency differences of their initial segments (words: high, non-words: low). (Note that this is contradicted by the fact that responses on average were made at roughly 400 ms after word offset; nonetheless, it is still possible that listeners had made their *decisions* prior to word offset.) However, in the difficult condition, listeners had to attend to more of the word (the non-words were cohorts of real words) before having enough information to make a decision. This attention to later parts of the word – especially the parts that are most affected by talker variation, vowels (Bachorowski & Owren, 1999; Hertz, 2006; Owren & Cardillo, 2006) – may have been the operative factor in allowing talker effects to emerge.

McLennan and Luce (2005) also presented data from a speeded shadowing task (Experiment 3) in which, again, talker-specific effects emerged (faster shadowing times for same-talker repetitions of words vs. different-talker repetitions). In one condition, listeners repeated words immediately upon hearing them, and no talker effects were seen. In another condition, listener responses were delayed by 150 ms, and talker-specific effects emerged. Presumably, the additional processing time in the delay condition allowed talker information to interact with the lexical information and thus influence responses. This is at odds with predictions based on a system where talker information is located within the lexicon. It would be interesting to know if the spontaneous talker-imitation effects observed by Goldinger (1998) were present in McLennan and Luce's data, despite the lack of RT differences. However, the fact remains that no RT differences were noticed until relatively late for participants in their shadowing task, supporting the hypothesis that talker information has late and thus extralexical effects. These "late" results will be considered again in the General Discussion.

In an attempt to provide additional information about the time course with which talker-specific information is used in lexical processing, we conducted two experiments using paradigms and methods that have shown how segmental information is used to disambiguate temporarily competing lexical representations. All current models of spoken word recognition assume that as a word unfolds over time, partially activated lexical candidates compete for recognition (Gaskell & Marslen-Wilson, 1999; Norris, 1994; Marslen-Wilson, 1987; McClelland & Elman, 1986). Supporting evidence comes from demonstrations that recognition of a word is affected by the characteristics of words with similar sounds (e.g., Luce & Pisoni, 1998; Marslen-Wilson, 1987, 1990;

Vitevitch, 2002; Vitevitch & Luce, 1998). Some of the strongest evidence comes from studies demonstrating that words which share initial segments with a target word, so-called "cohorts", become temporarily activated as the input unfolds (e.g., Allopenna, Magnuson, & Tanenhaus, 1998; Magnuson, Tanenhaus, Aslin, & Dahan, 2003; Marslen-Wilson & Zwitserlood, 1989). Cohort competitors decrease in activation as inconsistent information accumulates. Mismatching information can be segmental, such as the difference between /kæp/ and /kæt/, which diverges at their final phoneme, or it can be a sub-phonemic mismatch, such as temporarily misleading coarticulatory information (Dahan, Magnuson, Tanenhaus, & Hogan, 2001; Marslen-Wilson & Warren, 1994; McQueen, Norris, & Cutler, 1999), or a mismatch in typical duration (Salverda, Dahan, & McQueen, 2003).

Arguably the most fine-grained information about the time course of lexical competition comes from word recognition studies using the visual world paradigm (Cooper, 1974; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995), in which the listener's gaze fixations to potential referents are monitored as the speech unfolds over time. For instance, Allopenna et al. (1998) found that listeners adjust nearly instantaneously to a divergence between *candy* and *candle*, showing more looks to *candy* 200 ms after the disambiguating segment (/i/ vs. /l/) is reached. (Two hundred milliseconds is the delay commonly assumed for the latency to plan and execute a saccadic eye movement, Hallett, 1986.) Magnuson et al. (2003) report similar results with an artificial lexicon, which allows tighter control over word frequency and word similarity.

In the present pair of experiments, we utilize the eye-tracking methodology to examine whether talker differences can be used by our participants, in this experimental context, as a cue to lexical identity. The first experiment examines the role of talker variation in disambiguating natural-language cohort pairs in a 4-alternative referent selection task. For some cohort pairs, both words are spoken by the same talker. For other cohort pairs, one word is spoken by a female talker and the other by a male talker, giving these pairs an additional distinguishing piece of information, at least in the context of the experiment. There are several blocks of presentations, and we look at the time course of disambiguation of same-talker and different-talker cohorts in the first two blocks – prior to the listener's exposure to many instances of voice-specific presentations – and in the last two blocks, after the listener has been exposed to nearly 20 presentations of each word.

The second experiment uses an artificial lexicon (Magnuson et al., 2003) to examine talker variability in a set of stimuli that are less English-like. Given that novice language learners (infants and L2 learners) often show the largest effects of talker variability, one reasonable hypothesis is that phonologically unfamiliar new words will produce even stronger talker effects than the more phonologically familiar ones in Experiment 1.

In summary, this research had three goals. First and foremost, we wished to determine whether talker variation, normally a non-phonemic acoustic attribute, could be used in lexical disambiguation. To this end, we set out to estimate the time course of such disambiguation, utilizing the eye-tracking methodology. Second, in Experiment 2, we examined the impact of informative talker information on lexical access of newly learned words in an artificial lexicon. Third, we examined in Experiment 2 how phonological similarity interacts with potential talker disambiguation effects, by examining talker effects for cohort and rhyme competitors.

2. Experiment 1

In this experiment, we asked whether or not listeners could implicitly use talker identity to disambiguate lexical items during spoken word recognition. As the strongest test, we focused on cohort words – known to cause large amounts of lexical competition – and consistently assigned one member of some cohort pairs (*talker-mismatched*) to one of two different voices, such that talker was a cue to lexical identity for those words (e.g., a male voice saying *sheep* and a female voice saying *sheet*). Other cohort pairs (*talker-matched*) were consistently spoken by the same talker (male *couch*, male *cows*) so that, for these words, talker was not a cue to lexical identity. In the final set of trials after exposure to these talker-lexical item mappings, we looked for evidence of faster disambiguation for the talker-mismatched pairs.

2.1. Method

2.1.1. Participants

Sixteen University of Rochester undergraduates participated in this experiment and were paid \$10 for a 1 h and 10 min session. All were native English speakers and reported no history of hearing problems.

2.1.2. Stimuli

Stimuli were 48 consonant–vowel–consonant (CVC) words drawn from the Kucera and Francis database (1967), using the Irvine Phonotactic Online Dictionary (Version 1.2, Vaden & Hickok, 2004). Words deemed to be non-imageable were eliminated. Words that ended in sonorant consonants or liquids (m, n, ŋ, r, l) were also eliminated from the set, as these consonants have more influence on the quality of the vowel preceding them than do other consonants (J.M. McDonough, personal communication). We selected 12 cohort pairs (24 words) and 24 filler words. Both cohort and filler words fell below the criterion for "highly frequent" as defined by Goldinger (1998). Our reasoning, following Goldinger's (1998), was that words which occur very frequently would have too many traces to be substantially influenced by the word-talker contingencies we presented during the course of our relatively brief experiment. The sets were split into two subgroups that were assigned to be same-talker or different-talker for equal numbers of participants. Word subgroups (Appendix A) did not differ in terms of frequency (raw, $p \geqslant .85$; log-weighted, $p \geqslant .44$) or neighborhood density (raw, $p \geqslant .07$, but $p > .3$ between cohort groups; frequency-weighted, $p \geqslant .34$; log-weighted, $p \geqslant .1$, but $p > .6$ between cohort groups).

All 48 words were recorded in a sound-attenuated chamber by both a female native speaker of English and a male native speaker of English. The male talker was born in Rochester, Minnesota, and attended public schools there until the age of 18. Following this, he attended college in Connecticut for 4 years, and lived in the Czech Republic (1 year) and Minneapolis (3 years) before attending graduate school. The female talker was born near Columbia, South Carolina, and attended public schools and an undergraduate university there. Both talkers were completing graduate studies in western New York at the time of the recordings, where they had resided for the previous 6 years.

Several acoustic attributes distinguish the talkers. Most notable is an approximately octave-sized disparity in F_0 (90 Hz vs. 190 Hz). Other potential differences are speaking rate, F_0 variability, and vowel quality. The exact nature of the vowels of each talker likely differ, as they were raised in different regions of the United States, and vowel variation is commonplace among English dialects. For the present experiment, the male talker had a mean F_0 of 97 Hz, the female talker an F_0 of 184 Hz, a significant difference by a paired *t*-test ($t(47) = 10.09$, $p < .0001$). Words for the female talker averaged 561 ms in duration, and for the male talker 559 ms, which was not a significant difference ($t(47) = .20$, $p = .84$). As the words were monosyllabic, pitch variation across syllables was not measured (though see Experiment 2).

We should note that there has been some debate that words may be encoded by talker gender and not talker identity (the connotative hypothesis of Geiselman & Crawley, 1983). That is, priming effects associated with talker identity are caused by affective connections with male vs. female voices, and not by perceptual specificity. However, Palmeri, Goldinger, and Pisoni (1993) found that priming occurs for a talker even when other talkers in the set are of the same gender, and priming does *not* occur across different talkers of the same gender, supporting a more perceptual basis for talker effects.

Words were listed in multiple random orders, and each talker read through each list. Recordings were made in a sound-attenuated chamber to a PC running Praat (Boersma, 2003) through a CSL language processing box at a sampling rate of 22,050 Hz, and were then written to .wav files. These files were edited in SoundEdit 16 software to select the clearest token of each word for each talker.

The 12 pairs of cohort words were subdivided into two sets, A and B. Filler words were also assigned to A or B subgroups. For half of the listeners, each set-A cohort pair was talker-matched; that is, both words were spoken by the same talker throughout the experiment. Three pairs were selected to be spoken by the female talker, the other three by the male. This was randomly assigned for each participant. For set B, each cohort pair was talker-mismatched, such that one word was always spoken by the female talker, and the other was always spoken by the male talker. Again, the talker-word assignment was randomized for each participant. For the other half of the listeners, the reverse was true: set-B words were talker-matched and set-A words were talker-mismatched. For all listeners, the non-cohort words were pronounced equally often by either talker, serving to disguise the talker-specificity hypothesis.

Picture referents for each word were selected from www.clipart.com. Pictures were chosen to be similar in style and clarity. We made an effort to select the most representative pictures possible, but given our limited stimulus set, some of the items did not have prototypical picturable references (e.g., "dish"). Therefore, prior to the experiment, participants were shown pictures of the items with orthographic labels beneath. (Orthographic labels were used to avoid presenting the words in a particular voice prior to the experiment proper.) This process of seeing the picture of each word along with its orthographic label took 2–5 min and was done to ensure reasonable familiarity with the pictures used.

2.1.3. Procedure

The experimental stimuli were presented in custom software on a PC running Windows XP. The experimental computer was networked to a second PC that controlled an EyeLink II head-mounted eye tracker and recorded gaze position every 4 ms. The software recorded the screen coordinates, stimulus presentations, and participant responses for analysis in a database program. Software provided by the EyeLink system was used to detect saccades using the thresholds for motion (0.2°), velocity (30°/s) and acceleration (8000°/s^2).

Each word was presented 20 times over the course of 960 trials, which were split into ten 96-trial blocks. Within a block, each cohort word was presented twice, with all occurrences of a given word presented in the same voice. Each non-cohort word was also presented twice, spoken once by the female talker and once by the male talker. The reason for blocking in sets of 96 instead of 48 trials was to allow for equivalent numbers of presentations of non-cohort items by each talker within a block.

At the beginning of each trial, a central fixation point appeared, and participants fixated this point while clicking the mouse. This provided drift correction information to the eye tracker and triggered the simultaneous presentation of four pictures. After 500 ms, the name of the target picture was spoken. Participants then clicked on one of the four pictures. The design of the program did not allow them to continue to the next trial until they had clicked the correct picture. After selecting the correct response, all four pictures disappeared and the fixation point reappeared. Recalibration of the eye tracker was (rarely) performed between trials as needed.

For each participant, the same four pictures always appeared together (see Fig. 1), but in different spatial locations across trials. This was done to prevent the expectation that certain *pairs* of pictures preferentially appeared together and that, if they did, the target was likely to be one of them. That is, given a fixed set of four pictures, any one of them was equally likely to be the target word spoken by one of the two talkers. When a word with a cohort was a target, its competitor's picture was always present in the same 4-picture display. Non-cohort words were equally likely to be the spoken targets. We analyzed only the trials where the cohort words (not the filler words) were targets.

2.2. Results and discussion

For each 4 ms time slice on each trial, the location of gaze was categorized. If the screen coordinates of the gaze were in one of the four pictures or in a surrounding

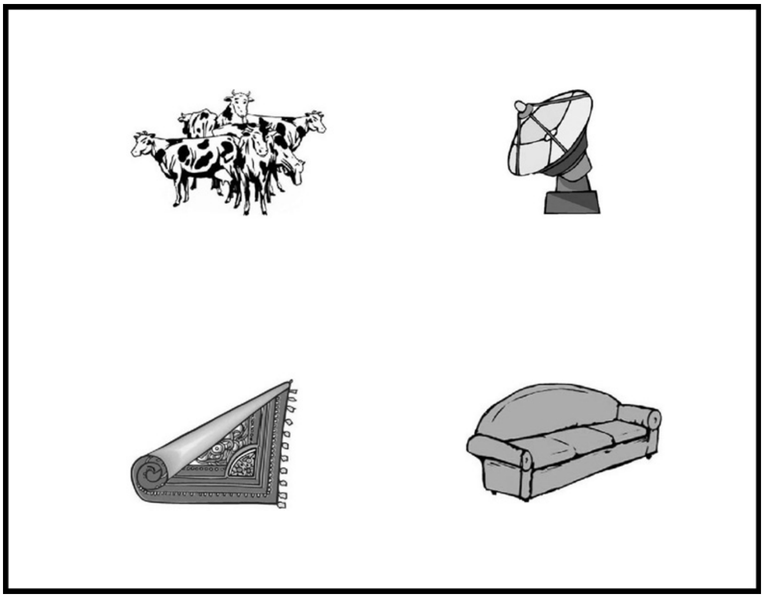


Fig. 1. An example trial from Experiment 1. On each trial, the participants heard a word (e.g., "cows") and were instructed to mouse-click the appropriate referent picture. A cohort competitor (*couch*) was also present.

200-pixel area, gaze was classified as being on that picture. Looks to any other areas of the screen, including the center, were classified as "no looks to objects." An additional category specified that there was no reliable fixation data at that time point. Later processing categorized looks to a given picture in a given position on a given trial (e.g., fixating the sheep in the upper left corner) according to the variables of interest (e.g., looks to all talker-mismatched targets). A typical trial began with time slices in the "no looks" category because participants were fixating the central point (not displayed on the graphs). Thus, when looks are summed across trials, the proportions of looks to particular objects (target, cohort, distractor) all rise from 0 to some asymptotic level less than 1 because of the no-looks category. For computational tractability, we averaged fixation proportions into 100-ms bins (0–96 ms, 100–196 ms, etc.).

We analyzed the cohort trials (i.e., when either member of a cohort pair was spoken) from the first two blocks (first 20% of trials) and the cohort trials from the last two blocks (last 20% of trials). Fig. 2 presents fixation proportions to the target words, cohort competitors and unrelated distractors, for the 2000 ms after the onset of the target word for the first (Fig. 2a) and last (Fig. 2b) two blocks, respectively. The results replicate standard cohort effects, with fixations to the target and cohort increasing together (and exceeding the distractors) until about 600 ms, and then separating as disambiguating information arrives. We focused on a 200–800 ms time window following word onset, which corresponds roughly to the point where the first

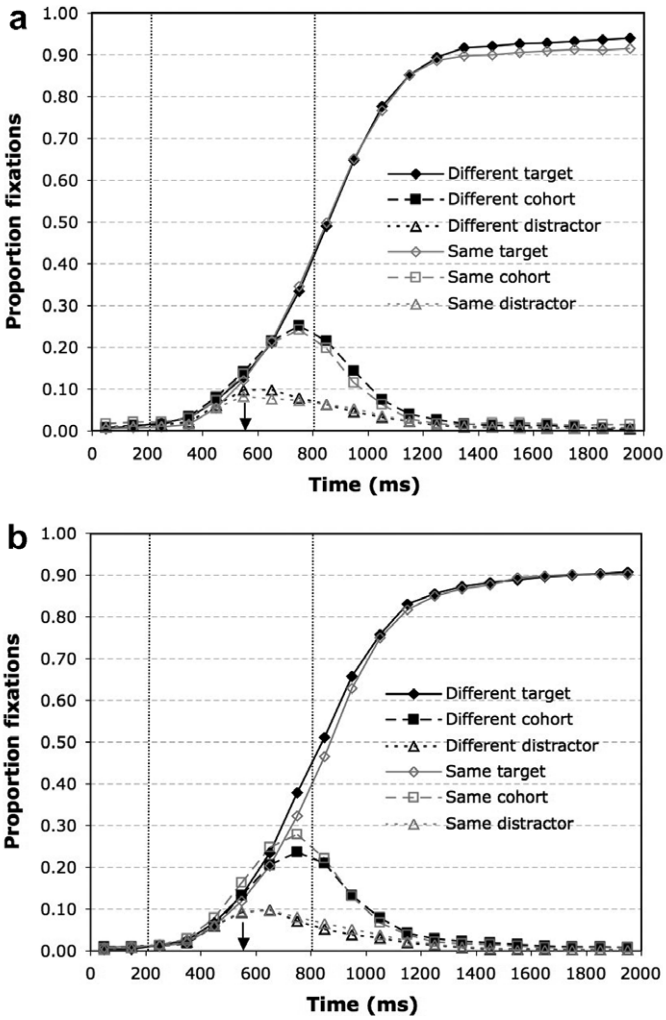


Fig. 2. Fixations to targets, cohort competitors, and distractors in (a) the first 20% of trials (pre-talker exposure) and (b) the last 20% of trials (post-talker exposure) in Experiment 1. Each point represents the center of a 100-ms bin (e.g., 0–96 ms or 500–596 ms). Arrow indicates average word offset, dotted lines indicate the time window analyzed.

signal-driven fixations are expected, to roughly 200 ms after the offset of the target words, which averaged 560 ms in length. In the first two blocks (Figs. 2a and 3), looks to same-talker cohort competitors did not differ from looks to different-talker competitors. In the last two blocks (Figs. 2b and 3), looks to same-talker competitors exceeded looks to different-talker competitors.

These results were confirmed statistically by a repeated-measures analysis of variance (ANOVA) on competitor fixations, with Block (first two, last two) and Talker (different talkers, same talker) as factors.[1,2] There was a main effect of Talker ($F_1(1,15) = 6.22$, $p = .02$, $F_2(1,23) = 3.22$, $p = .09$, $\eta^2_G = .010$),[3] such that overall there tended to be a smaller proportion of fixations to different-talker competitors than to same-talker competitors. There was no effect of Block ($F_1 < 1$, $F_2 < 1$). Most importantly, there was a reliable Block × Talker interaction ($F_1(1,15) = 5.5$, $p = .03$, $F_2(1,23) = 5.33$, $p = .03$, $\eta^2_G = .015$), such that the difference in fixation proportions between different and same talkers, which did not differ in the first two blocks, was reliable in the last two blocks (first two: $t_1(15) = .52$, $p = .61$, $t_2(23) = .5$, $p = .62$, mean difference = .004, Cohen's $d = .066$; last two: $t_1(15) = 4.16$, $p = .0008$, $t_2(23) = 3.13$, $p = .005$, mean difference = −.025, $d = .435$).[4]

The results clearly demonstrated a talker disambiguation effect. After repeated presentations of cohort pairs in a single voice or two different voices, competitor activation for same-voice cohort pairs was greater than for different-voice cohort pairs. Moreover, this effect was relatively early as the word was unfolding; for the last two blocks the proportion of fixations to same and different cohorts began to diverge about 500 ms after the onset of the word, indicating that the talker effects on fixations had begun before the offset of the word.

These findings support, but are distinct from, previous experiments (Goldinger, 1996, 1998; Palmeri et al., 1993) in demonstrating a direct effect of talker information on word recognition – a reduction in competition effects with different talkers, rather than an advantage for repetitions of the same word by the same talker, or for talker-homogeneous presentation of lists. The results are somewhat counter to McLennan and Luce's (2005) findings of only late talker effects. It is not entirely clear how best to compare the timing of manual or vocal responses in their tasks to the saccadic eye

[1] A similar analysis of targets showed no significant effects. A combined analysis of target advantage scores (target fixations minus competitor fixations) yielded results qualitatively similar to the competitor-only analysis with the exception that the interaction of Block and Talker was not significant (importantly, the two target advantage scores diverged significantly only in the final 20% of trials), but because results are not at all evident in the targets and strongly evident in the competitors, we have reported only the competitor analysis.

[2] Raaijmakers, Schrijnemakers, and Gremmen (1999) and Raaijmakers (2003) show convincingly that the *F*-test by participants is the appropriate unbiased estimator of treatment effects in counterbalanced designs. We therefore make statistical decisions and conclusions based on F_1, while reporting F_2 values by convention. We thank an anonymous reviewer for drawing this to our attention.

[3] The decision of an effect size measure for a repeated-measures design is not a trivial one, as there has been much debate as to whether or not to include variability due to participants. For ANOVA effects, we report generalized eta-squared (η^2_G) as defined by Olejnik and Algina (2003) and as recommended for within-subjects designs by Bakeman (2005). This measure ranges from 0 to 1 and describes the variance accounted for by a factor in a similar way to partial eta-squared (η^2_P), but includes variability due to participants in the error term. This makes it numerically smaller than η^2_P, but more generalizable across designs. For *t*-tests, we report Cohen's *d* (Cohen, 1988), which is a standardized mean difference. We calculated these values from the means and standard deviations of each distribution, which includes individual participant variability to prevent overestimating effect size (see Dunlap, Cortina, Vaslow, & Burke, 1996). Unlike η^2_G, the absolute value of the *d* measure can vary from 0 to $\pm\infty$ as the standard deviation approaches 0, but in practical terms Cohen (1988) provides reference values of .2, .5, and .8 as small, medium, and large effect sizes. All effect sizes reported are calculated from by-participants analyses.

[4] Note that this is roughly a "medium" effect size according to Cohen (1988).

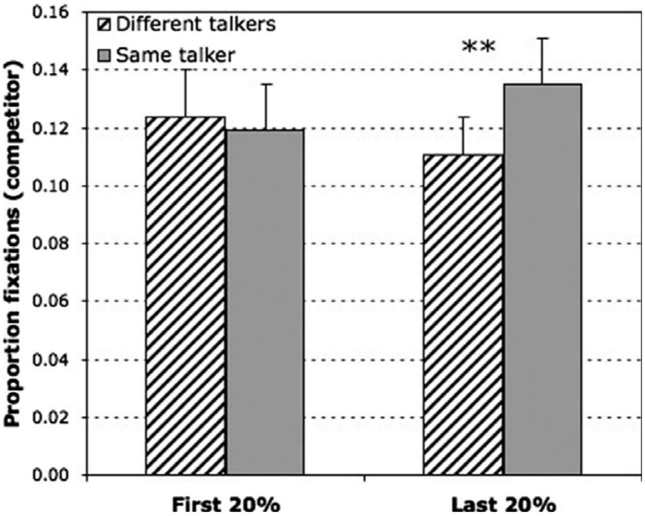


Fig. 3. Fixations to competitors in the first and last 20% of trials in Experiment 1, 200–800 ms. Error bars are standard errors.

movement timing in our task, given that manual responses are points in time and our fixation proportions are calculated across a time window. Their manual RTs were around 800 ms for stimuli that were 373 ms in duration on average. Given 300 ms to make a manual response, responses were made at 500 ms after onset, and thus 100 ms after offset. If one takes the first visible divergence of competitor fixations as the comparable response time (say, the 500–600 ms bin), our results are indeed "fast" relative to their manual reaction times. In fact, if one assumes that it takes 200 ms to program a signal-based saccade, this response is emerging based on the first 300–400 ms of the words (which were 560 ms in duration on average), i.e., 160–260 ms *prior* to word offset.

3. Experiment 2

The results of the first experiment provided support for incorporation of talker information into the lexicon, consistent with a view that the lexicon consists of multiple overlying episodic traces (Goldinger, 1996, 1998). Interference from cohort competitors (*sheep*, *sheet*) is attenuated as early as 200–800 ms after word onset when talker information disambiguates these words. However, the effects were modest, affected only the competitor, and were presumably limited by a lifetime of multi-talker exposure to these words. Perhaps low-frequency words, had there been a sufficient available imageable set, would have shown greater effects, but even low-frequency words might show small talker-specific effects if they were composed of common phonological patterns. Thus, Experiment 1 does not reveal a role for talker information in acquiring unfamiliar phonological patterns (either in L1 or L2;

Bradlow et al., 1997; Houston & Jusczyk, 2000, 2003; Lively et al., 1993; Logan et al., 1991). The current experiment explores this issue.

Additionally, certain portions of the speech signal may differ more between talkers than other portions. For example, vowels, which carry fundamental frequency and other aspects of the glottal waveform, may be more robust correlates of talker variability than consonants (Bachorowski & Owren, 1999; Hertz, 2006; Owren & Cardillo, 2006). Unfamiliar (new) words with talker information that is available in the beginning of the word may demonstrate talker-based disambiguation even more clearly. We examine this in the following experiment using an artificial lexicon (e.g., Magnuson et al., 2003) of words with the desired properties that would be difficult to achieve with natural language stimuli. An artificial lexicon with the visual world paradigm also allows us to contrast an explicit measure of talker effects – error rates (see Creel, Aslin, & Tanenhaus, 2006) – with an implicit measure, namely gaze fixations to referents of newly learned words (see Magnuson et al., 2003).

Magnuson et al. (2003) presented participants with 16 consonant–vowel–consonant–vowel (CVCV) words. Each word had two neighbors within the novel lexicon, as defined by the Shortcut Rule in Luce's Neighborhood Activation Model (NAM, Luce & Pisoni, 1998); that is, they differed by the addition, deletion, or substitution of a single segment. The words were subdivided into sets of four, such that each word (e.g., *pibo*) had a cohort neighbor (*pibu*) and a rhyme neighbor (*dibo*). The fourth word in the set differed from the target word by two segments (*dibu*) such that the other two words (*pibu*, *dibo*) served as its rhyme and cohort neighbors, respectively. There was also a frequency manipulation. Half of the words occurred with high frequency at training, and half at low frequency, in a 7:1 ratio. Within low- and high-frequency words, the cohort and rhyme neighbors could also be lower or higher in frequency than the target, yielding four combinations of frequency and type of competitor.

At test, each word was presented as a target with pictures of a neighbor (either a cohort or a rhyme) and two distractors also visible in the display. Magnuson et al. (2003) found that there was more competition overall from cohort neighbors than from rhyme neighbors, and that these rhyme effects diminished across the two days of the experiment. In addition, activation of the targets was modulated by both target and competitor frequency: high-frequency words were activated more rapidly than low-frequency words, and words with high frequency competitors (neighbors) were activated more slowly. Importantly, processing of words in the artificial lexicon was at best only minimally affected by the neighborhood structure of the native (English) lexicon.

In the present experiment, we adapted this design to examine how talker differences affected lexical activation. Rather than varying frequency, we varied talker match among competitors: female-talker word with male-talker competitors, female-talker word with female-talker competitors, and vice versa. The data of primary interest were possible mismatch effects between words with neighbors (lexical competitors) that were talker-matched (spoken by the same talker), and words that had talker-mismatched neighbors (spoken by different talkers).

3.1. Method

3.1.1. Participants

A total of 19 participants were drawn from the same pool as in Experiment 1. None had participated previously in any similar experiments.

3.1.2. Stimuli

Novel words were recorded by the same two talkers as in Experiment 1, using the same equipment and software. Again, the clearest instances of each word were selected, and tokens with obvious deviations (e.g., in amplitude, length, vowel quality, or nasality) from the other stimuli were eliminated. Over the course of 75- to 90-min sessions on each of two days, the participants were exposed to 16 artificial vocabulary items repeated in blocks across multiple trials. These items were labels for 16 unfamiliar black-and-white shapes (Appendix B) that were a subset of pictures created for an earlier study (Creel et al., 2006). The novel lexicon used here mirrored that of Magnuson et al. (2003), Experiment 1, except that instead of varying word frequency, we varied talker (same or different from the word's competitor). Each set of words was spoken either by the same talker, or by two different talkers, with assignment to conditions counterbalanced across participants. The pictures used were uniquely randomly assigned to words for each participant.

Unlike the preceding experiment, the words in this experiment (Table 1) were VCVCV in structure, with primary stress on the second syllable. These changes served two functions. First, the stimuli were made less English-like than their counterparts in Experiment 1, decreasing the likelihood that episodic traces of them would be swamped by existing similar phonological traces from English. (Note that, as non-words, they were already extremely unlikely to have *lexical* traces that could interact with existing lexical items.) Second, all Cs were liquids or approximants (r, l, w, j), in order to have the most vowel-like words possible. Only two VCVCV words using these consonants appear in IPhOD for English (Vaden & Hickok, 2004): *aurora* and *arroyo*. Given that numerous talker-variable characteristics (F_0, F_0 variability, length) are carried by vowels, extremely vowel-like words may be more likely to show talker effects. Words differed in terms of F_0 (male: 86 Hz; female: 189 Hz; $t(15) = 52.12$, $p < .0001$), F_0 variability (male: 1.17; female: 1.75; $t(11) = 8.66$, $p < .0001$),[5] and length (male: 790 ms; female: 932 ms; $t(15) = 10.8$, $p < .0001$).

3.1.3. Procedure

The experiment was conducted over two days. Each word occurred six times in each of seven blocks on each of two days, for a total of 84 presentations. On each day, participants received training trials and testing trials under the control of PsyScope software (Cohen, MacWhinney, Flatt, & Provost, 1993), with the exception of the test on the second day, which was run using the custom software

[5] This was calculated as syllable $2F_0$/syllable $3F_0$, but the effect holds for syllable $2F_0$/syllable $1F_0$ as well. Some samples were omitted from this statistical test because the pitch tracking algorithm failed in the male talker's syllable 3 productions due to creaky voice.

Table 1
Stimuli for Experiment 2

aruja	**aruju**	**eɪwala**	eɪwaleɪ
eɪruja	**eɪruju**	owala	**owaleɪ**
ujoweɪ	ujowo	oleɪro	oleɪru
ajoweɪ	**ajowo**	uleɪro	uleɪru

Note. Plain typeface indicates male talker, boldface indicates female talker. Assignments of words to talkers were rotated through the participants.

described earlier to allow tracking of participants' visual fixations. Training trials occurred over a number of blocks per day, following Magnuson et al. (2003). On the first day, the first three blocks were 2AFC, and the last four were 4AFC. On the second day, all seven blocks were 4AFC. There were thus 672 training trials per day.

On each training trial within each block, the participant clicked on a cross at the center of the screen, which triggered the presentation of two (or four) pictures (see Fig. 4). After 500 ms, a name was spoken in isolation. Participants clicked on a picture, at which point the incorrect alternative(s) disappeared, leaving only the correct referent. The name of the referent was spoken again, and then the next trial proceeded. For 2AFC cases (the first three blocks on Day 1; see Fig. 4a), objects on each trial appeared randomly as 200×200 pixel pictures in two of four screen locations (upper left [384, 256], upper right [896, 256], lower left [384, 768], lower right [896, 768] on a monitor set at a resolution of 1280×1024). Screen locations for 4AFC trials were also randomized, as was the order of trials within each block for each participant.

The test (see Fig. 4b) was identical to the 4AFC exposure phase, except that participants received no feedback. There were 64 unique test trials, with each new word occurring 4 times as a target: once with a cohort competitor present, once with a rhyme competitor present, and twice on filler trials with no phonologically related competitors present. These 64 trials occurred twice (once in each of two blocks) for a total of 128 test trials per day. As before, trial orders and picture positions within a trial were randomly determined. After a response was made, correct or incorrect, the next trial began. Errors were recorded throughout training and testing. Eye movements were recorded during the second test. Each session of training and testing lasted approximately 75–90 min.

3.2. Results and discussion

Three participants were eliminated from the dataset due to experimenter error or equipment difficulties. Two participants were eliminated because their error rates exceeded 15% on the second day of testing. An additional participant was eliminated for consistent failure to make eye movements, instead fixating the central point on most trials. The final sample consisted of 13 participants.

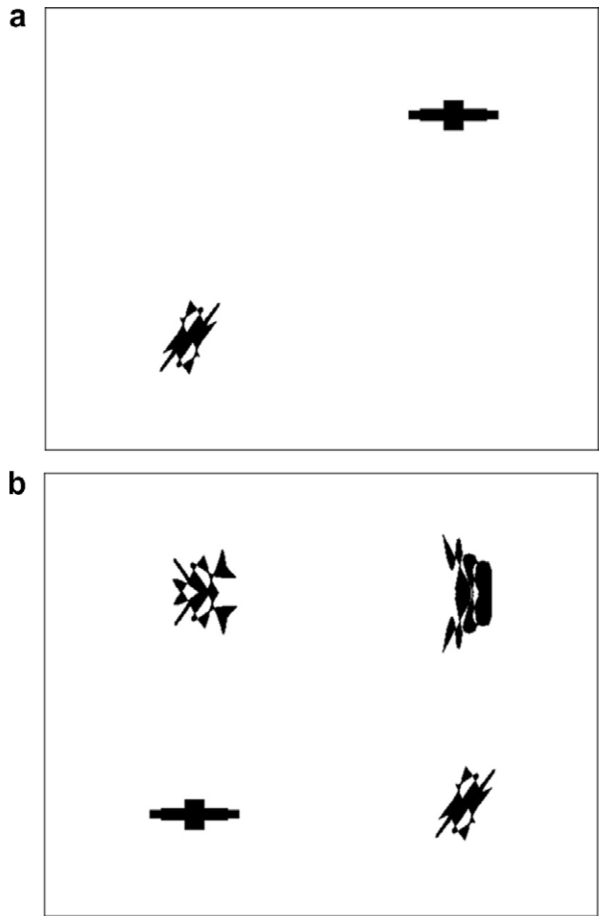


Fig. 4. Examples of displays on (a) 2AFC training trials and (b) 4AFC training and test trials in Experiment 2.

3.2.1. *Test errors*

We examined participants' performances on the test (with no feedback) for each day (Fig. 5). The error data (mouse-clicking the incorrect picture) were restricted to competitor-present trials. Specifically, we compared percent errors to competitors (cohorts or rhymes) to percent errors to unrelated distractors. As there were two distractors and only one competitor present on each trial, the distractor scores were averaged in order to equate chance performance. That is, by chance, selection of a distractor from the three non-target choices would occur twice as often as selection of the competitor. A significant difference between competitor and distractor error rates indicates that listeners find the target more lexically confusable with the competitor than with the distractors. This analysis is similar to that applied to fixation proportions as in Experiment 1, as well as to the analysis of error data in other work from our laboratory (Creel, Tanenhaus, & Aslin, 2006; Creel et al., 2006).

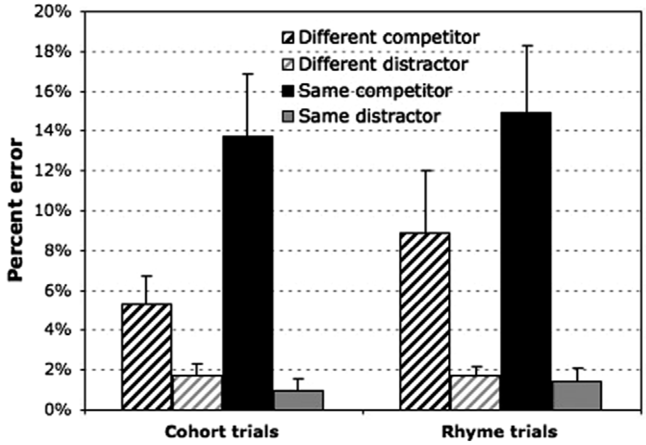


Fig. 5. Test errors in Experiment 2, collapsed across days. Error bars are standard errors.

Error rates decreased over days. There were more errors to phonologically related items (competitors) than to phonologically unrelated items, and there were fewer errors to different-talker competitors than to same-talker competitors. Errors on cohort and rhyme trials were roughly equivalent. We confirmed these results with a repeated-measures ANOVA with Day (first, second), Talker (different, same), Competitor Type (cohort, rhyme), and Error Type (competitor, distractor) as factors. There were decreases in errors from the first to the second test (effect of Day, $F_1(1,12) = 11.14$, $p = .006$, $F_2(1,15) = 15.02$. $p = .002$, $\eta^2_G = .024$). As expected, there was an effect of Error Type (F_1 $(1,12) = 40.6$, $p < .0001$, $F_2(1,15) = 84.3$, $p < .0001$, $\eta^2_G = .237$), with more competitor than distractor errors. This competitor–distractor asymmetry decreased from the first to the second day (Day × Error Type interaction, $F_1(1,12) = 6.53$, $p = .03$, $F_2(1,15) = 5.54$, $p = .03$, $\eta^2_G = .011$), with a more precipitous decrease in competitor errors than distractor errors. There was no difference in error rates for cohorts and rhymes (Competitor Type effect, $F_1(1,12) = .57$, $p = .47$, $F_2(1,15) = 2.81$, $p = .11$, $\eta^2_G = .006$). Finally, there was an effect of Talker ($F_1(1,12) = 6.55$, $p = .03$, $F_2(1,15) = 7.02$, $p = .02$, $\eta^2_G = .040$), with fewer errors for words with different-talker competitors. This effect was carried primarily by fewer competitor errors (Talker × Error Type interaction, $F_1(1,12) = 9.73$, $p = .009$, $F_2(1,15) = 14.38$, $p = .002$, $\eta^2_G = .051$) on the different-talker trials than the same-talker trials ($t_1(12) = 2.9$, $p = .01$, $t_2(15) = 3.25$, $p = .005$, mean difference = .072, $d = .965$), with no difference in distractor errors ($t(12) = .85$, $p = .4$, $t_2(15) = .79$, $p = .44$, mean difference = .005, $d = .249$). No other effects or interactions approached significance.

3.2.2. Eye movement data

We analyzed eye movements only on those trials where the correct referent was selected, to minimize spurious competitor fixations based on actual competitor selec-

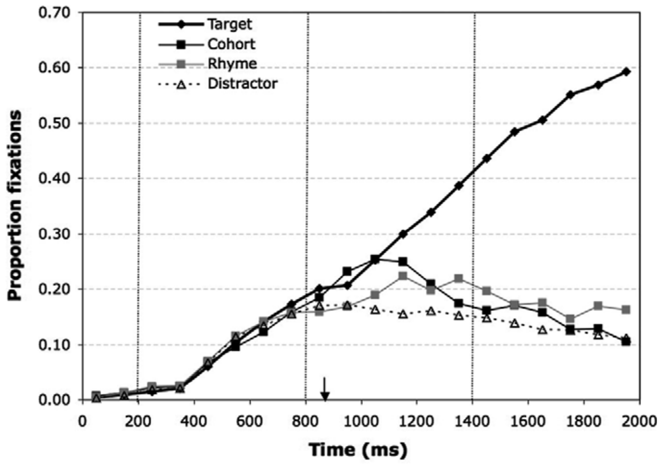


Fig. 6. Overall fixation proportions to targets, cohorts, rhymes, and distractors. Each point represents the center of a 100-ms bin. Arrow indicates average word offset, dotted lines indicate the time window analyzed.

tion (see Magnuson et al., 2003). In all other respects, processing of eye movement data was identical to that in Experiment 1.

Overall fixation patterns (Fig. 6) were similar to those seen in Experiment 1 and other experiments finding competitor effects (e.g., Allopenna et al., 1998; Magnuson et al., 2003), except that the patterns were relatively slower. This tends to be true of both longer words and relatively new words (see Magnuson et al., 2003). Fixations to the target and cohort increase together, passing distractor looks around 800 ms, with rhyme looks increasing above the distractor baseline slightly later, replicating the pattern found by Allopenna et al. and Magnuson et al. Thus, while fixation proportions are slower relative to Experiment 1, they show orderly incremental processing of the signal. Fixation proportions over time are subdivided into different-talker and same-talker trials in Fig. 7.

Since target activation as indexed by fixation proportions was slow relative to the previous experiment, we subdivided the time period from 200 to 2000 ms into three 600-ms windows of analysis (Fig. 8). In the 200–800 ms window used in Experiment 1, looks to all potential referents (including distractors) were indistinguishable; effects of segmental overlap and talker occurred in later time windows, with effects of talker beginning at about the same time as targets and cohorts began to diverge from the distractors. In contrast to Experiment 1, there were clear effects of talker on fixations to the targets, with fewer fixations to targets with same-talker competitors compared to targets with different-talker competitors. Similar to Experiment 1, fixations to competitors were lower for different-talker competitors than same-talker competitors, but this effect did not emerge until the last time window.

For cohort and rhyme trials, we analyzed targets and competitors in separate repeated-measures ANOVAs with Talker (different, same) and Time Window

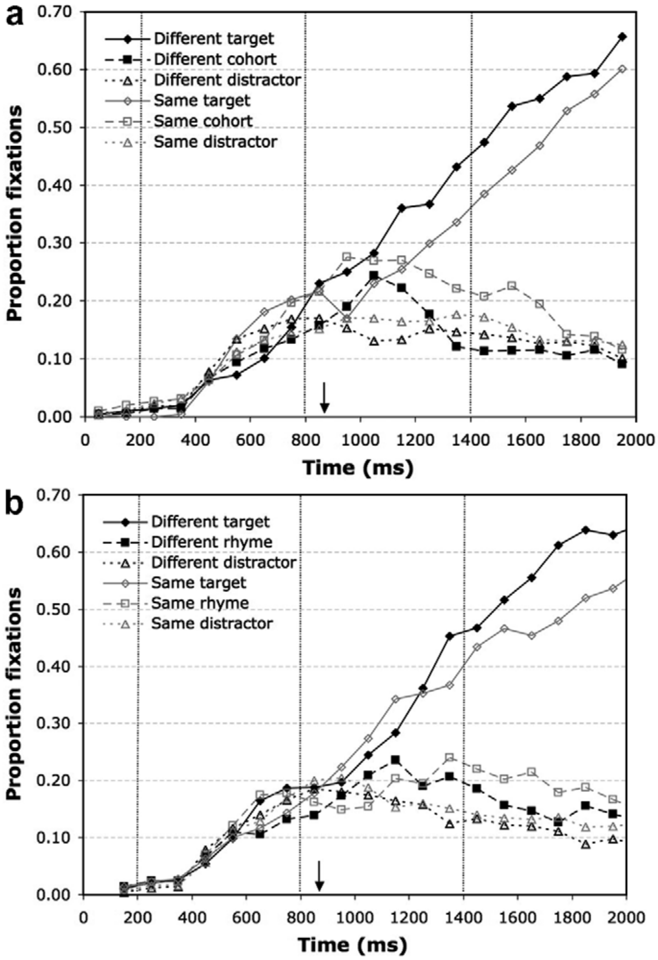


Fig. 7. Experiment 2 fixation data for (a) cohort trials and (b) rhyme trials. Each point represents the center of a 100-ms bin. Arrow indicates average word offset, dotted lines indicate the time windows analyzed.

(200–800, 800–1400, 1400–2000) as factors.[6] For targets on cohort trials, there was a marginal effect of Talker, ($F_1(1,12) = 4.19$, $p = .06$; $F_2(1,15) = 2.29$, $p = .15$,

[6] As in Experiment 1, a combined analysis of target advantage scores for cohort trials (target fixations minus cohort fixations) yielded results much like the target-only and competitor-only analyses. The Talker × Object Fixated interaction did not reach significance until the third window. However, it is worth noting that the target–cohort difference achieved significance in the second window (800–1400 ms) for different-talker trials (second window: $t_1(12) = 3.24$, $p = .007$, $t_2(15) = 2.64$, $p = .02$, $m = .115$, $d = 1.17$; third window: $t_1(12) = 7.93$, $p < .0001$, $t_2(15) = 8.16$, $p < .0001$, $m = .47$, $d = 3.69$). The same test only achieved significance in same-talker trials by the third (1400–2000 ms) window (second window: $t_1(12) = .35$, $p = .73$, $t_2(15) = 2.03$, $p = .06$, $m = .012$, $d = .13$; third window: $t_1(12) = 5.43$, $p = .0002$, $t_2(15) = 5.76$, $p < .0001$, $m = .31$, $d = 2.68$). Thus, while activation does seem to be slower in Experiment 2 than Experiment 3, a talker difference still confers an advantage.

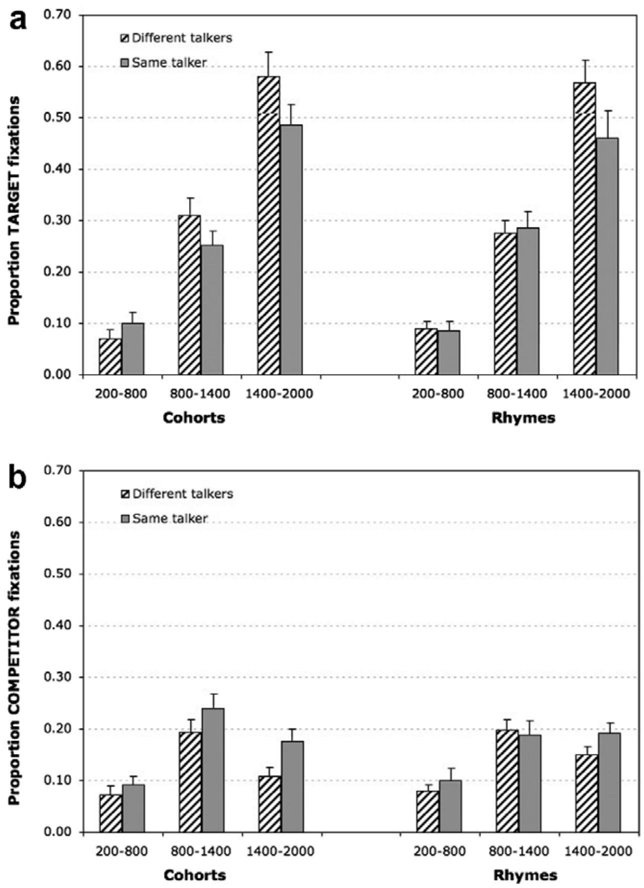


Fig. 8. Experiment 2 fixation data in cohort and rhyme trials in designated windows of analysis. (a) Targets and (b) competitors. Error bars are standard errors.

$\eta^2_G = .032$), but there was an effect of Time Window ($F_1(2,24) = 79.473$, $p < .0001$; $F_2(2,30) = 268.4$, $p < .0001$, $\eta^2_G = .730$) and a Talker × Time Window interaction ($F_1(2,24) = 7.51$, $p = .003$; $F_2(2,30) = 3.84$, $p = .03$, $\eta^2_G = .050$). The effect of Time Window indicates the expected rise in fixation proportions over time. The interaction with Talker indicated that target fixation proportions for different talker targets surpassed those for same-talker targets by the last time window (first window: $t_1(12) = -1.77$, $p = .1$, $t_2(15) = 1.92$, $p = .07$, $m = -.029$, $d = .42$; second window, $t_1(12) = 1.65$, $p = .12$, $t_2(15) = 1.97$, $p = .07$, $m = .058$, $d = .55$; third window: $t_1(12) = 3.44$, $p = .005$, $t_2(15) = 1.59$, $p = .13$, $m = .093$, $d = .59$).

For targets on rhyme trials, a similar but nonsignificant pattern occurred. Again, there was no main effect of Talker ($F_1(1,12) = 1.49$, $p = .25$; $F_2(1,15) = 3.57$, $p = .08$, $\eta^2_G = .022$) but a strong effect of Time Window ($F_1(2,24) = 84.09$, $p < .0001$; $F_2(2,30) = 193.2$, $p < .0001$, $\eta^2_G = .707$), and a Talker × Time Window interaction that was marginal ($F_1(2,24) = 3.38$, $p = .051$; $F_2(2,30) = 2.00$, $p = .15$, $\eta^2_G = .050$). As on cohort trials, the Time Window effect indicated a rise in fixation proportions to targets over time. The marginal interaction with Talker resulted from a marginal advantage for different-talker targets in the final time window ($t_1(12) = 1.94$, $p = .08$; $t_2(15) = 2.21$, $p = .04$, $m = .106$, $d = .63$).

For competitor fixations on cohort trials, the results were somewhat parallel to, but not identical to, the target analyses. This time, there was an effect of Talker ($F_1(1,12) = 8.87$, $p = .01$; $F_2(1,15) = 18.27$, $p = .0007$, $\eta^2_G = .086$), indicating an overall lower proportion of fixations to different-talker cohort competitors than same-talker cohorts ($m = -.044$, $d = .933$). An effect of Time Window ($F_1(2,24) = 16.61$, $p < .0001$; $F_2(2,30) = 40.58$, $p < .0001$, $\eta^2_G = .374$) suggested a rise in competitor fixations from the first to the second time window for both different-talker and same-talker cohorts ($t_1(12) = 6.72$, $p < .0001$; $t_2(15) = 8.37$, $p < .0001$, $m = .135$, $d = 2.12$), but a decrease from the second to the third time window ($t_1(12) = 2.75$, $p = .02$; $t_2(15) = 4.6$, $p = .0003$, $m = .076$, $d = 1.13$). The interaction with Talker was not significant ($F_1(2,24) = 1.32$, $p = .29$; $F_2(2,30) = 1.35$, $p = .28$, $\eta^2_G = .020$), though it is worth noting that the Talker effect (talker-matched cohort vs. talker-mismatched cohort) was only significant in the final time window (first window: $t_1(12) = -1.32$, $p = .21$, $t_2(15) = -.97$, $p = .35$, $m = -.018$, $d = .304$; second window: $t_1(12) = -1.69$, $p = .12$, $t_2(15) = -3.64$, $p = .002$, $m = -.045$, $d = .517$; third window: $t_1(12) = -2.73$, $p = .02$, $t_2(15) = -2.84$, $p = .01$, $m = -.067$, $d = .897$).

For competitor fixations on rhyme trials, however, the only significant effect was that of Time Window ($F_1(2,24) = 15.48$, $p < .0001$; $F_2(2,30) = 17.06$, $p < .0001$, $\eta^2_G = .294$), suggesting a rise from the first to second time window in competitor fixations ($t_1(12) = 7.42$, $p < .0001$; $t_2(15) = 7.85$, $p < .0001$, $m = .104$, $d = 1.99$), though not from the second to the third window ($t_1(12) = 1.1$, $p = .29$; $t_2(15) = 1.05$, $p = .31$, $m = -.022$, $d = .456$). The effect of Talker ($F_1(1,12) = 1.39$, $p = .26$; $F_2(1,15) = 2.02$, $p = .18$, $\eta^2_G = .013$) was not significant, nor was the Talker × Time Window interaction ($F_1(2,24) < 1$; $F_2(2,30) < 1$).

These results provide additional evidence for an effect of talker-specific disambiguation, with a set of newly learned stimuli that are vastly different from those that were used in Experiment 1. A direct comparison is inadvisable due to the different stimulus lengths and growth in fixation proportion curves, but we can tentatively say that the effect here is more robust. For comparable measurements, consider the effect sizes of the talker effect for cohort competitors in the final block of Experiment 1 ($d = .435$) vs. the final-time-window talker effect for cohort competitors ($d = .897$) or cohort targets ($d = .59$) in the current experiment. There are numerous reasons why this might be so. First, as mentioned earlier, the words in Experiment 2 bore limited resemblance to English words, and therefore the talker-specific traces we presented carried more influence than those in Experiment 1. Second, cohort disambiguation based on vowels may well be slower because vowels are apprehended more slowly than consonants due to their greater dialectal and indexical variability (see Cutler, van Ooijen, Norris, & Sanchez-Casas, 1996), implying that there is more time for talker information to have an effect.

However, if this were true, rhymes (differing only by vowels initially) should also have been affected by talker early on, yet they were not affected as strongly as cohorts.

It is still plausible, however, that the phonology of these newly learned VCVCV words is sufficiently low-probability in English that it causes very slow lexical access. One argument might be that the talker effects shown here are late (cf., McLennan & Luce, 2005), and that as the phonology of the words becomes overlearned, talker information would be unnecessary and fade away. The competing prediction of exemplar models is that talker effects would only become more robust and reliable with time and exposure. It would be interesting to present these stimuli for longer periods of time to determine the extent to which greater familiarity leads to faster activation, and to the maintenance (or disappearance) of talker effects on lexical access.

Another interesting point is that talker effects for rhymes were not prominent in the fixation data, despite the rather high frequency of rhyme errors overall. However, this unpredicted effect may be due to the presence of segmental information that disambiguates words early, leading to minimal talker effects (the rhyme competitor case). By contrast, when segmental information is *not* present to disambiguate words, there are large talker effects (the cohort competitor case). It is also worth noting that there are hints of a rhyme talker effect in the final 1400–2000 ms analysis window, though this is not significant; such a late effect is consistent with the late overlap (relative to cohorts) of rhymes with each other. The later the overlap, the later talker effects can come into play.

4. General discussion

We had several goals in the present work, which were met to varying degrees. First, we asked whether a non-phonemic acoustic attribute – talker variation – could be used in lexical disambiguation. The answer is a definitive yes: in both experiments we found fewer competitor fixations or more target fixations, or both, for different-talker cohort pairs than for same-talker pairs. This suggests that different-talker lexical items compete less with each other than same-talker lexical items do. In Experiment 1, the degree of this disambiguation was not as strong as that based on segmental cues, but in the face of a lifetime of exposure and experience that does not support talker specificity in lexical representation, our experimental manipulation was still effective.

More in depth, we sought to pinpoint the time course of talker-based disambiguation, utilizing the head-mounted eye-tracking methodology. We found in Experiment 1 that for overlearned words that were recently correlated with a specific talker, talker-based disambiguation can occur as early as 500 ms (roughly) after the onset of a word. In Experiment 2 for newly learned words, this disambiguation was somewhat slower. Crucially, however, talker effects were not delayed relative to phonemic effects. It is likely that more training on our artificial lexicons would result in faster target activation, but precisely how such faster activation would modify the talker disambiguation effect is unclear. With respect to McLennan and Luce's (2005) findings, Experiment 1 suggests that talker-specific effects can occur relatively early. One reason for this may be that talker was a relevant cue in our task (it was an additional cue to word identity), but not in their lexical decision or shadowing tasks. Experiment 2 is more in line with their results, with talker effects emerging relatively late. One might ascribe these results to slow integration of talker information, or to the imperfect learning of the lexicon overall rather than the slowness of talker information *per se*. Given the differences in technique (priming and shadowing vs. gaze fixations and error rates) and numbers of exposures (2 vs. 40 or more, counting learning trials), it is difficult to select between these alternatives. Future research is needed to determine how the relatively slow effects in our Experiment 2 might evolve as word recognition becomes more rapid: they might dissipate, or, alternatively, they might remain as rapid as phoneme-based recognition processes.

Second, we explored the impact of informative talker variation on lexical access of newly learned words. The motivation for using an artificial lexicon was that we could create precisely controlled acoustic/phonetic structures: identical word and phoneme frequencies within the set, and specific phonological (and talker) competitor relationships among words. This allowed us to create a situation that is more akin to learning a first language or a phonologically unfamiliar second language. It is in these situations that one might expect stronger effects of talker variation (or other sorts of acoustic variation) for unknown words, and this is certainly borne out for infants (Barker & Newman, 2004; Houston & Jusczyk, 2000, 2003) and L2 learners (Bradlow et al., 1997; Lively et al., 1993; Logan et al., 1991). Our work suggests that talker information does affect the encoding of new words. The error data from Experiment 2 imply an overall facilitative effect of differentiating competitors by talker, but one might argue that attention to talker is a deliberate strategy used by participants. Countering this argument, the eye movement data we obtained are more convincing as an implicit index of talker effects, and, tellingly, these differential effects based on talker only occur when phonemic information is not present to disambiguate words (i.e., when the competitors are cohorts).

Our final goal was to examine the effects of phonological resemblance to English on the magnitude of the talker effect. Because the experiments varied on multiple parameters (English or novel words, consonant-initial or vowel-initial, familiar or unfamiliar phonological patterns), it is not possible to make more than tentative statements on this matter. However, in addition to the already-mentioned larger effect sizes in Experiment 2 than in Experiment 1, some other pieces of information are suggestive. There was some evidence of phonological familiarity, in that English stimuli (Experiment 1) manifested a different-talker competition advantage only as weaker competitor effect, while less English-like stimuli (VCVCVs in Experiment 2) evidenced effects for both target fixations and cohort fixations. This may be to some extent due to imperfect learning in Experiment 2 – with better learning, target fixations might rise to asymptotic levels. However, it is likely that competitor fixations would still differ more strongly than those in Experiment 1: note that competitor fixations for different-talker cohorts in Experiment 2 are quite close to distractor fixations (see Fig. 7a), while in Experiment 1, there were still noticeable fixations to different-talker cohorts above the distractor baseline. As noted before, the difference in the Experiment 2 effect might be due also to the highly vocalic nature of the stimuli, not just their dissimilarity from English words (though their vowel-heavy nature is one of the features that makes them less English-like). Perhaps words are less well-identified by vowels than by consonants (cf., Creel et al., 2006). Nonetheless, vowel-differentiated minimal pairs (*us*, *ice*, *ace*, *S*) are clearly identifiable by their vowels (cf., Bonatti, Peña, Nespor, & Mehler, 2005, for evidence that vowels are less important than consonants in word segmentation), and we saw clear vowel-based disambiguation in the rhyme-competitor trials of Experiment 2. The fact that *any* sort of word is sensitive to talker information suggests that talker information may indeed be stored in the lexicon.

A final interpretive issue concerns whether non-phonemic talker-specific information is part of the lexical representation, *per se*, or whether it is an example of paired-associate learning or context-dependent memory that naturally falls out of correlated cues in the input. One could ask whether we would see the same effects if we embedded consistent contextual sounds with word pairs (e.g., "couch" always spoken with a car horn in the background, "cows" always spoken with a slide whistle in the background), or even if we associated pictures with melodies that began identically but differed in timbre.[7] Is talker information effective because it is part of the speech signal and is explicitly represented in the lexicon, or because it is an associated cue?

The associated-cue experiments described above might well produce similar results to those we have reported here for talker variation in the lexicon. Such an outcome – and perhaps our own data – would lend credence to a distributed view of the lexicon (e.g., Elman, 2004), rather than a view in which the sound representations of words are limited to only those types of variation that are contrastive in some natural language. According to the more traditional view, the sound representation of words in the lexicon is limited to only those types of variation that are contrastive in some natural language. Higher-level lexical representations are limited to combinatory lexical information that affects distributional constraints on how words combine with one another, such as argument structure (Koenig, Mauner, & Bienvenue, 2003; Tanenhaus & Carlson, 1989) and perhaps more enriched event representations that participate in generative processes (e.g., Pustejovsky, 1991). In contrast, according to the distributed view, a lexical entry is a coactivated set of representations that encompass all properties of a word, such as the word's phonological form, motor actions (Hauk, Johnsrude, & Pulvermüller, 2004), visual speech information (McGurk & MacDonald, 1976), and combinatory information such as verb argument structure (e.g., MacDonald, Pearlmutter, & Seidenberg, 1994; Tanenhaus, Carlson, & Trueswell, 1989). These properties are correlated with one another, and will upon presentation coactivate each other.

[7] We thank Arthur Samuel and an anonymous reviewer for raising this issue, and Arthur Samuel for suggesting these thought experiments.

In some cases, such coactivation can be quite compelling (see McGurk & MacDonald, 1976).

Evidence from functional brain imaging supports the notion of distributed representations (e.g., Hauk et al., 2004; Martin, Wiggs, Ungerleider, & Haxby, 1996). Recently, Hauk et al. had participants passively read verbs for actions executed by different body parts (*lick*, *pick*, *kick*; tongue, hand, foot). They found that the three verb types (tongue-, hand-, and foot-words) elicited differential activation in bilateral primary motor cortex, varying according to the body part used. This suggests that lexical semantic representations for these verbs have distributed patterns of activation that differ according to actual motoric functions associated with those verbs. Moreover, there seems to be no principled way to cordon off these motor representations from the rest of the "lexical" information associated with these words. Thus, lexical entries may include similarly distributed types of information. If talker information, car horns, and slide whistles correlate with certain words, then these pieces of information will be coactivated with all of the other parts of the word's representation. Rather than dividing information related to a word into "lexical" and "extralexical," all information associated with a word may be part of its lexical entry.

The foregoing account is also consistent with the intuition that melodies differing in timbre paired with pictures would elicit similar results: what learners are acquiring is the correlation between the word/melody and the referent picture. Rather than positing completely separate systems for word-learning and association-learning, a common learning mechanism can parsimoniously account for both processes.

How do our results inform this debate? A reasonable criterion would be rapidity – if the multiple aspects of a lexical entry are coactivated, then the influence of one aspect (talker) on the others should be reasonably instantaneous. In our Experiment 1, the use of talker information was, while not large in magnitude, relatively rapid, making plausible the notion that talker information was activated simultaneously with other aspects of the word's lexical entry. Experiment 2, with greater magnitudes of talker effects, did not show the same rapidity. Talker effects did, however, emerge at about the same point in time as phonemic effects. Thus, our results provide some support for the distributed view of lexical representation by showing some degree of association of talker information with other aspects of lexical representations.

In addition, the current data can serve as a benchmark for future work on other correlated cues to word identity. For instance, how rapidly – in terms of the amount of exposure needed – can one learn to utilize novel acoustic cues to word identity that, unlike talker variability, are external to the speech signal (e.g., car horns and slide whistles)? While we agree that the "car horn" manipulation would likely be effective, it is possible that there may be some priority for incorporating signal-internal (or auditory-source-internal; Bregman, 1990) information, such that information that comes from a discernibly different auditory source than the speech signal (e.g., a car horn) shows less robust effects. On the other hand, the greater novelty and acoustic dissimilarity of car horns and slide whistles, relative to talker differences, might give these external sounds *greater* cue strength. (An additional possibility is that auditory grouping mechanisms themselves are learned from very robust correlational information, but such a discussion is beyond the scope of this paper.)

In conclusion, the present pair of experiments demonstrated facilitation of lexical disambiguation by talker information, both for familiar and for newly learned words. Effects occurred relatively early in the time course of lexical access, at least for familiar words, as indexed by gaze fixations during a referent selection task. For a novel lexicon, talker information affected both off-line confusions and transient on-line lexical competition. The transition from effects on novel lexicons to effects on the mature lexicon remains a topic for future investigation.

Acknowledgments

Thanks to Joyce McDonough for many valuable conversations with regard to this research, and to Arthur Samuel, Steve Goldinger, and an anonymous reviewer for valuable comments on this manuscript. Thanks also to Owen Brimijoin for patiently recording seemingly unending lists of words. This research was supported by a NSF Graduate Research Fellowship to SCC, by NIDCD Grant DC-005071 to M.K.T. and R.N.A., and by Packard Foundation Grant 2001-17783 to R.N.A. These studies were conducted as part of the first author's dissertation.

Appendix A

Frequency and neighborhood information, Experiment 1 stimuli (KF refers to Kucera & Francis (1967) frequency counts).

Word	Neighborhood density	Frequency-weighted density	Log density	KF frequency	Log frequency
Cohort pairs					
Set A					
Bag	28	1683	24.19	42	1.62
Bat	42	22,396	53.49	18	1.26
Couch	9	128	7.17	13	1.11
Cows	14	244	10.34	16	1.20
Knife	9	1594	13.87	80	1.90
Knight	38	7736	51.86	21	1.32
Map	32	1584	21.75	13	1.11
Match	25	2466	20.86	42	1.62
Pop	34	614	26.11	8	0.90
Pot	46	7053	48.59	28	1.45
Sheep	29	3778	32	24	1.38
Sheet	34	4144	40.98	47	1.67

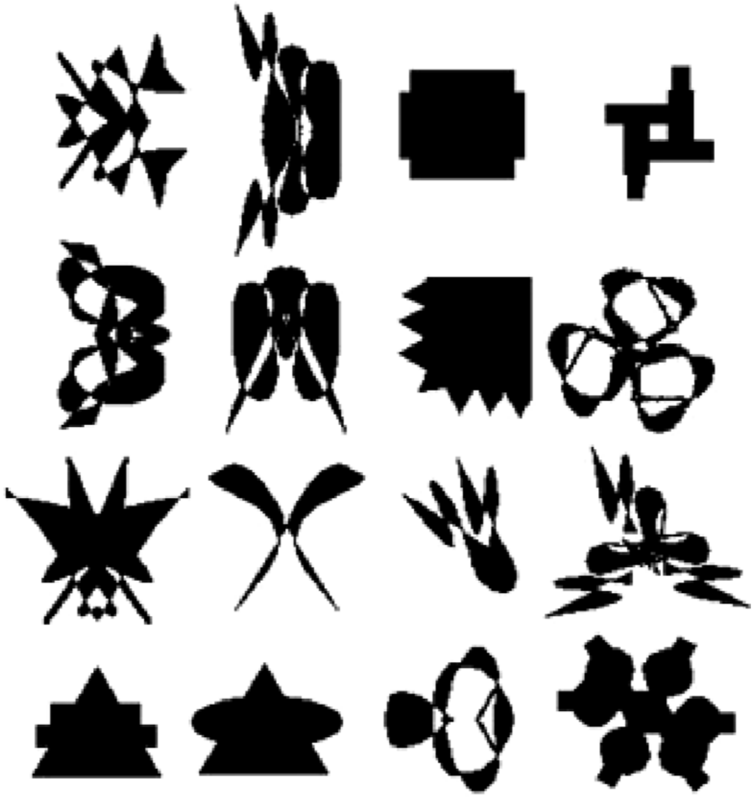
(*continued on next page*)

Appendix A (*continued*)

Word	Neighborhood density	Frequency-weighted density	Log density	KF frequency	Log frequency
Set B					
Boot	36	4974	32.34	14	1.15
Booth	16	1010	16.48	8	0.90
Cage	22	1595	24.58	9	0.95
Cake	34	2861	36.96	13	1.11
Maid	47	3249	46.5	31	1.49
Maze	51	4500	47.39	6	0.78
Nose	37	4767	38.3	60	1.78
Note	34	8978	40.46	127	2.10
Pies	39	1096	35.01	5	0.70
Pipe	19	450	17.54	20	1.30
Soup	28	403	16.34	16	1.20
Suit	31	1320	31.3	49	1.69
Distractors					
Set A					
Dish	17	1351	16.94	16	1.20
Rug	27	439	17.02	14	1.15
Goat	27	2602	27.67	6	0.78
Juice	14	923	15.33	12	1.08
Toes	42	1597	34.01	19	1.28
Fish	12	534	13.94	35	1.54
Hose	51	12,593	57.98	9	0.95
Duck	39	1131	28.79	9	0.95
Foot	13	1320	17.64	113	2.05
Sub	15	3325	18.01	21	1.32
Leg	21	1306	20.24	62	1.79
Soap	27	2451	21.37	23	1.36
Set B					
Shirt	19	394	14.74	28	1.45
Leaf	24	1328	23.06	15	1.18
Lock	48	2995	45.45	25	1.40
Witch	32	10,064	29.55	5	0.70
Hat	43	28,824	55.38	57	1.76
Tub	16	275	13.88	13	1.11
Toys	19	273	11.61	11	1.04
Ship	27	438	23.4	84	1.92
Hook	19	1182	17.24	5	0.70
Teeth	14	400	14.76	103	2.01
Wreath	19	572	16.33	8	0.90
Log	13	1311	16.51	14	1.15

Appendix B

Visual stimuli used in Experiment 2.



References

Allen, J. S., Miller, J. L., & DeSteno, D. (2002). Individual talker differences in voice-onset-time. *Journal of the Acoustical Society of America, 113*, 544–552.

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory & Language, 38*, 419–439.

Andruski, J., Blumstein, S., & Burton, M. (1994). The effect of subphonetic differences on lexical access. *Cognition, 52*, 163–167.

Bachorowski, J.-A., & Owren, M. J. (1999). Acoustic correlates of talker sex and individual talker identity are present in a short vowel segment produced in running speech. *Journal of the Acoustical Society of America, 106*, 1054–1063.

Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior Research Methods, Instruments, and Computers, 37*, 379–384.

Barker, B. A., & Newman, R. S. (2004). Listen to your mother! The role of talker familiarity in infant streaming. *Cognition, 94*, B45–B53.

Bonatti, L. L., Peña, M., Nespor, M., & Mehler, J. (2005). Linguistic constraints on statistical computations: The role of consonants and vowels in continuous speech processing. *Psychological Science, 16*, 451–459.

Bradlow, A. R., Pisoni, D. B., Yamada, R. A., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/ IV: Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America, 101*, 2299–2310.

Bradlow, A. R., Torretta, G. M., & Pisoni, D. B. (1996). Intelligibility of normal speech I: Global and fine-grained acoustic–phonetic talker characteristics. *Speech Communication, 20*, 255–272.

Bregman, A. S. (1990). *Auditory scene analysis*. Cambridge: MIT.

Clopper, C. G., & Pisoni, D. B. (2004). Effects of talker variability on perceptual learning of dialects. *Language & Speech, 47*, 207–239.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavior Research Methods, Instruments & Computers, 25*, 257–271.

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language. A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology, 6*, 84–107.

Creel, S. C., Aslin, R. N., & Tanenhaus, M. K. (2006). Acquiring an artificial lexicon: Segment type and order information in early lexical entries. *Journal of Memory & Language, 54*, 1–19.

Creel, S. C., Tanenhaus, M. T., & Aslin, R. N. (2006). Consequences of lexical stress on learning an artificial lexicon. *Journal of Experimental Psychology: Learning, Memory & Cognition, 32*, 15–32.

Cutler, A., van Ooijen, B., Norris, D., & Sanchez-Casas, R. (1996). Speeded detection of vowels: A cross-linguistic study. *Perception & Psychophysics, 58*, 807–822.

Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes, 16*, 507–534.

Dunlap, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods, 1*, 170–177.

Eisner, F., & McQueen, J. M. (2006a). The specificity of perceptual learning in speech processing. *Perception & Psychophysics, 67*, 224–238.

Eisner, F., & McQueen, J. M. (2006b). Perceptual learning in speech: Stability over time (L). *Journal of the Acoustical Society of America, 119*, 1950–1953.

Elman, J. (2004). An alternative view of the mental lexicon. *Trends in Cognitive Sciences, 8*, 301–306.

Gaskell, M. G., & Marslen-Wilson, W. D. (1999). Ambiguity, competition, and blending in spoken word recognition. *Cognitive Science, 23*, 439–462.

Geiselman, R. E., & Crawley, J. M. (1983). Incidental processing of speaker characteristics: Voice as connotative information. *Journal of Verbal Learning & Verbal Behavior, 22*, 15–23.

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

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105*, 251–279.

Hallett, P. E. (1986). Eye movements. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (pp. 10.1–10.112). New York: Wiley.

Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron, 41*, 301–307.

Hertz, S. (2006). A model of the regularities underlying speaker variation: Evidence from hybrid synthesis. In *Proceedings of the ninth international conference on spoken language processing, Pittsburgh, PA*.

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review, 95*, 528–551.

Houston, D. M., & Jusczyk, P. W. (2000). The role of talker-specific information in word segmentation by infants. *Journal of Experimental Psychology: Human Perception & Performance, 26*, 1570–1582.

Houston, D. M., & Jusczyk, P. W. (2003). Infants' long-term memory for the sound patterns of words and voices. *Journal of Experimental Psychology: Human Perception & Performance, 29*, 1143–1154.

Koenig, J.-P., Mauner, G., & Bienvenue, B. (2003). Arguments for adjuncts. *Cognition, 89*, 67–103.

Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal?. *Cognitive Psychology 51*, 141–178.

Kraljic, T., & Samuel, A. G. (2006). How general is perceptual learning for speech?. *Psychonomic Bulletin & Review 13*, 262–268.

Kucera, H., & Francis, W. (1967). *Computational analysis of present day American English*. Providence, RI: Brown University Press.

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/: II. The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America, 94*, 1242–1255.

Logan, J. D., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America, 89*, 874–886.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear & Hearing, 19*, 1–36.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review, 101*, 676–703.

Magnuson, J. S., Tanenhaus, M. K., Aslin, R. N., & Dahan, D. (2003). The time course of spoken word learning and recognition: Studies with artificial lexicons. *Journal of Experimental Psychology: General, 132*, 202–227.

Marslen-Wilson, W. (1987). Functional parallelism in spoken word recognition. *Cognition, 25*, 71–102.

Marslen-Wilson, W. (1990). Activation, competition, and frequency in lexical access. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives ACL–MIT Press series in natural language processing* (pp. 148–172). Cambridge: MIT Press.

Marslen-Wilson, W., & Warren, P. (1994). Levels of perceptual representation and process in lexical access: Words, phonemes, and features. *Psychological Review, 101*, 653–675.

Marslen-Wilson, W., & Zwitserlood, P. (1989). Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human Perception & Performance, 15*, 576–585.

Martin, A., Wiggs, C. L., Ungerleider, L. G., & Haxby, J. V. (1996). Neural correlates of category-specific knowledge. *Nature, 379*, 649–652.

Maye, J., Werker, J., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82*, B101–B111.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18*, 1–86.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*, 746–748.

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

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition, 86*, B33–B42.

McQueen, J. M., Norris, D., & Cutler, A. (1999). Lexical influence in phonetic decision making: Evidence from subcategorical mismatches. *Journal of Experimental Psychology: Human Perception & Performance, 25*, 1363–1389.

Mirman, D., McClelland, J. L., & Holt, L. L. (2006). An interactive Hebbian account of lexically guided tuning of speech perception. *Psychonomic Bulletin & Review, 13*, 958–965.

Newman, R. S., Clouse, S. A., & Burnham, J. L. (2001). The perceptual consequences of within-talker variability in fricative production. *Journal of the Acoustical Society of America, 109*, 1181–1196.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition, 52*, 189–234.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology, 47*, 204–238.

Olejnik, S., & Algina, J. (2003). Generalized eta and omega squared statistics: Measures of effect size for some common research designs. *Psychological Methods, 8*, 434–447.

Owren, M. J., & Cardillo, G. C. (2006). The relative roles of vowels and consonants in discriminating talker identity versus word meaning. *Journal of the Acoustical Society of America, 119*, 1727–1739.

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

Pustejovsky, J. (1991). The generative lexicon. *Computational Linguistics, 17*, 409–441.

Raaijmakers, J. G. W. (2003). A further look at the "Language-as-Fixed-Effect Fallacy". *Canadian Journal of Experimental Psychology, 57*, 141–151.

Raaijmakers, J. G. W., Schrijnemakers, J. M. C., & Gremmen, F. (1999). How to deal with "The Language-as-Fixed-Effect Fallacy": Common misconceptions and alternative solutions. *Journal of Memory and Language, 41*, 416–426.

Salverda, A. P., Dahan, D., & McQueen, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition, 90*, 51–89.

Strange, W., & Dittmann, S. (1984). Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception & Psychophysics, 36*, 131–145.

Tanenhaus, M. K., & Carlson, G. N. (1989). Lexical structure and language comprehension. In W. D. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 529–561). Cambridge, MA: MIT Press.

Tanenhaus, M. K., Carlson, G. N., & Trueswell, J. C. (1989). The role of thematic structures in interpretation and parsing. *Language and Cognitive Processes, 4*, 211–234.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*, 1632–1634.

Vaden, K., & Hickok, G. (2004). The Irvine phonotactic online database (version 1.2) [on-line database]. Available from http://www.iphod.com/.

Van Lancker, D., Kreiman, J., & Wickens, T. D. (1985). Familiar voice recognition: Patterns and parameters, part II: Recognition of rate-altered voices. *Journal of Phonetics, 13*, 39–52.

Vitevitch, M. S. (2002). Influence of onset density on spoken-word recognition. *Journal of Experimental Psychology: Human Perception & Performance, 28*, 270–278.

Vitevitch, M. S., & Luce, P. A. (1998). When words compete: Levels of processing in perception of spoken words. *Psychological Science, 9*, 325–329.