# Talker information is not normalized in fluent speech: Evidence from on-line processing of spoken words

**Sarah C. Creel (creel@cogsci.ucsd.edu)**
University of California, San Diego, Department of Cognitive Science
La Jolla, CA 92093-0515

**Melanie A. Tumlin (mtumlin@cogsci.ucsd.edu)**
University of California, San Diego, Department of Cognitive Science
La Jolla, CA 92093-0515

## Abstract

Recent work demonstrates that talker characteristics can be used as predictive cues for spoken word recognition. However, abstractionist accounts suggest that talker information is usually stripped away or "normalized" based on preceding speech material. A contrasting account is that listeners never normalize, instead storing detailed, acoustically-varied instances of words. These varied instances then facilitate recognition of words in a vast variety of voices and accents. We present data suggesting that such "irrelevant" acoustic characteristics of word forms (talker-varying acoustic attributes) are not normalized, but are instead encoded, when learned in fluent speech context. Experiment 1 replicates recent demonstrations of talker specificity in word recognition. Using the same set of words in carrier sentences, Experiment 2 finds that learners still encode talker information even though in principle they could easily normalize away talker-based acoustic variability.

**Keywords:** talker variability; spoken word recognition; eye tracking; normalization; exemplar theory

## Introduction

One central question in spoken language processing has been how we recognize nonidentical or novel instances of known linguistic categories. The problem faced by a listener is that word forms exhibit a large amount of acoustic variability, but only a subset of that acoustic variability actually discriminates one word from another. The remaining variability in the speech signal has typically been regarded as useless noise. On this account, the presumed goal of the human language processing system is to remove irrelevant variation, leaving only that variability which cues word identity. Problematically, there is a different set of parameters to filter out depending on the utterance. For instance, fundamental frequency (F0) and accent characteristics can vary from talker to talker, and rate of speech can vary even within a talker (Van Lancker, Kreiman, & Emmorey, 1985; Van Lancker, Kreiman, & Wickens, 1985). Nonetheless, adult native speakers of a language are good at recognizing novel instances of a familiar word, and seem robust to acoustic variability among talkers.

Much research has been devoted to how listeners strain out or normalize over variability in the speech signal. Even if one assumes that humans are intrinsically equipped to detect particular speech sounds, listeners still must discern which subset of those speech sounds is relevant to word identity in her or his own language. Consistent with this, infants are more sensitive than adults to phonemically-irrelevant variability, including phoneme categories outside their own language (Werker & Tees, 1984), and acoustic variability among talkers (Houston & Jusczyk, 2000). Over development, the learner discovers that his or her spoken language input has certain dimensions (such as F0 variation, phoneme variation, rate variation), and further, that certain dimensions (phoneme variation, but not rate variation) are linked to changes in word meaning. Gradually, then, the learner would reweight attention toward phonemic variation and away from other types of variation. On this account, adult native speakers attend primarily to phonemic information, even in learning new words in their native language, as long as they are able to account for nonphonemic variation coming from another source (properties of the talker).

This approach contrasts with recent findings that talker information can be used to disambiguate words prior to their phonemic divergence point. Creel, Aslin, and Tanenhaus (2008) examined the time course of talker effects on word recognition by presenting listeners with known or novel words, where each word was perfectly correlated with a single talker. Creel et al. found that listeners showed less competition between different-talker pairs (male maid, female maze) than between same-talker pairs (male maid, male maze). More specifically, in an eye tracking paradigm, listeners hearing "maid" in a male voice showed fewer looks to the maze when it had consistently had a different talker than when it had the same talker as "maid." These data imply that adult listeners do not ignore talker information in learning and recognizing words.

Note here that talker information is not necessarily a conscious percept that there is a particular talker. By talker information, we refer to whatever set of acoustic attributes vary between talkers (such as F0, pitch variability, or speech rate, among other things). Of course, there are other types of acoustic variation that occur in speech as well, such as dialectal (accent) or idiolectal differences. In principle, listeners should be able to take advantage of any sort of acoustic variability (talker-related or otherwise) during word recognition, as long as that variability is sufficiently patterned. We have selected talker variability in particular because it interacts minimally with phonemic information (though see Allen, Miller, & DeSteno, 2003; Newman, Clouse, & Burnham, 2001). (This is much less true for accent variation, where the accent may alter particular phonemes to resemble other ones.)

The broader claim here is that the acoustic variability resulting from talker variation (or accent or idiosyncrasies of the talker) is intrinsic to representations of word form. On this view, the speech recognition system is not concerned with straining out talker variability, but instead with storing detailed information about how a particular word or speech sound varies as other attributes change (F0, for instance). There is suggestive evidence that listeners store talker variation along with word form information. For instance, listeners seem not to generalize certain phoneme shifts from one talker to another (Kraljic & Samuel, 2007), implying that at least some phonemic information may be stored in a talker-dependent fashion. Further, listeners learning a new second-language phoneme contrast are more successful at both perception and production when they hear exemplars from multiple talkers (Pisoni and colleagues; Logan, Lively & Pisoni, 1990; Lively, Logan & Pisoni, 1993; Lively, Pisoni, Yamada, Tokhura & Yamada, 1994; Bradlow, Pisoni, Akahane-Yamada & Tokhura, 1996). Goldinger (1996) demonstrated that correctly recognizing a previously-heard word as "old" improves as the test talker's perceptual similarity to the original talker increases.

However, it is as yet unclear how pervasive these talker-specificity effects are. It remains possible that, given time to identify talker-specific characteristics, adult listeners remove these characteristics from their representations. This is analogous to color perception: in isolation, a gray patch may appear to have a certain lightness, but given a surround of a particular lightness, viewers' perceptions automatically correct for the lightness level of the context. Here, we wanted to know if listeners correct for talker context in learning new words.

In Experiment 1, we replicated Creel et al.'s (2008) talker-specificity effect with a vocabulary of novel words. Having established this effect, we proceeded to investigate talker-specific learning of this vocabulary in sentence context in Experiment 2. To accomplish this, we utilized an artificial lexicon paradigm in combination with eye tracking. In this paradigm, listeners learn nonsense words as labels for unfamiliar objects. After learning, the listeners are asked to select one of the objects out of an array on a computer screen (Magnuson, Tanenhaus, Aslin & Dahan, 2003). Their eye movements are tracked as the spoken instruction (such as "Click on the X") unfolds over time. A high proportion of looks to an object indicates that listeners are considering that object to be a possible alternative—for instance, if listeners hear "sheep" and briefly visually fixate a picture of a sheet during "sheep," that would suggest that listeners were temporarily entertaining the hypothesis that the word they were hearing might be "sheet." In the current study, we had participants learn nonsense word labels where certain pairs of labels overlapped early, and either shared or did not share a talker.

## Experiment 1

Experiment 1 sought to replicate Creel et al. (2008) with a different artificial vocabulary than previously used. This new artificial vocabulary was very low in biphone frequency, but still roughly similar to common English words in the words' CV[C]C structure. We used this vocabulary instead of those from either of the two Creel et al. experiments for two reasons: the Creel et al. effects with real words were fairly subtle, and the effects with a highly English-atypical artificial lexicon were relatively late. By using this slightly easier-to-learn vocabulary, we hoped to find strong, rapid talker-specificity effects that could then be compared to a second experiment with the words learned in a sentential context.

### Method

**Participants.** Thirty-two native-English-speaking participants received course credit for taking part in this experiment.

**Stimuli.** We created a miniature lexicon of 16 nonsense monosyllabic words (Table 1). The 16 monosyllabic words comprised 8 pairs of words, where each pair matched until the final consonant. To lengthen slightly the time period during which each word was ambiguous, and to maximize the availability of fundamental frequency (F0) information, a correlate of talker identity, both initial and final consonants were voiced. This meant that F0 information was present from word onset, and that vowels were relatively long. (In English, vowels preceding a syllable-final voiced consonant are longer than those preceding voiceless consonants.)

Table 1: Word pairs used in Experiments 1 & 2.

| | |
|---|---|
| boog booj | belm beln |
| darg darj | dalm daln |
| veeg veej | vorm vorn |
| zelm zeln | zerg zerj |

One male and one female talker each recorded all 16 words, and tokens were selected on the basis of recording quality. Each word diverged from the other word in the pair (i.e., the final consonant began) around 362 milliseconds (ms) on average.

For each participant, in half of the pairs both words were spoken by the same talker (talker-same pairs), and for the other half of the pairs, each word was spoken by a different talker (talker-different pairs). Pair type (talker-same, talker-different) and talker (word spoken by male or female) were counterbalanced across words and participants.

Each word was learned as a label for one of 16 black-and-white pictures. These pictures (Figure 1) have been used in a number of previous word-learning experiments by Creel and colleagues. Shapes for similar-sounding word pairs were selected to be visually dissimilar, so as not to contaminate results with visual similarity effects. There were four different quasirandom assignments of words to pictures, to minimize the possibility of spurious similarity between words and pictures.

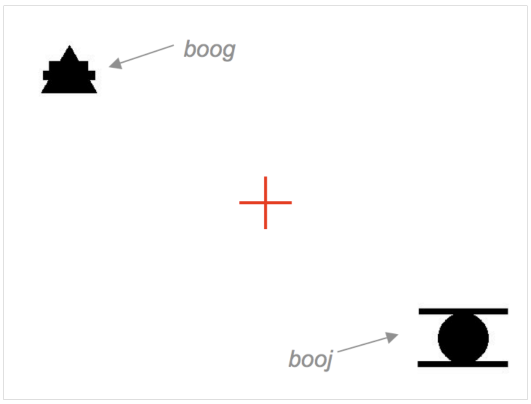


Figure 1. Sample learning trial. Labels for the two objects depicted are paired labels.

**Procedure.** Participants were presented with 128-trial blocks. Trials within a block were randomly ordered. On each trial, two pictures appeared and a label was spoken. The location of the pictures on each trial were counterbalanced to occur equally often at one of four positions, in the upper left, upper right, lower left, and lower right of the screen. The participant (initially guessing) mouse-clicked one picture as the one labeled, and received feedback in the form of the actually-correct picture remaining visible, while the incorrect shape disappeared. After reaching 90% correct on a 128-trial block, participants proceeded to the test phase, which was identical to training except that trials were not reinforced. At both training and test, on half the trials, the incorrect picture was the paired label (they heard "boog" and saw a boog and a booj). On the other half, the incorrect picture was a particular unpaired word (they heard "boog" and saw a boog and a daln). This controlled for effects of common presentation—if participants confuse words more simply because they are presented with the same pair of objects, then performance should be equally poor for paired-word trials and unpaired-word trials.

**Equipment.** During the experiment, participants' eye movements were monitored at a 2-ms sampling rate by an Eyelink Remote eyetracker (SR Research, Mississauga, ON). Custom Matlab software on a Mac Mini running OS 10.4 utilized PsychToolbox 3 (Brainard, 1997; Pelli, 1997) and the Eyelink Toolbox (Cornelissen, Peters & Palmer, 2002) to synchronize and communicate with the eyetracker. Data were processed offline using scripts written in Python to consolidate looking time data by participant and condition into 50-ms time bins.

## Results

**Accuracy.** Participants took two to six (m=3.3) 128-trial blocks to reach the 90% correct criterion. We performed an ANOVA with Competitor Type (paired, unpaired), Talker Match (same talker, different talkers) and Block (first training block, last training block, test) as within-participants factors. Performance is depicted in Figure 2. Accuracy increased over Block ($F(2,62) = 345.22$, $p < .0001$). Accuracy was much higher for unpaired phonemically dissimilar (boog, daln) trials than for paired trials (boog, booj; $F(1,31) = 143.18$, $p < .0001$). An interaction of Competitor Type and Block showed that the absolute difference in accuracy between phonemically dissimilar and phonemically paired trials decreased across blocks ($F(2,62) = 53.73$, $p < .0001$). A marginal interaction of Competitor Type x Talker Match ($F(1,31) = 3.91$, $p = .06$) reflected a nonsignificant advantage for different-talker items on paired trials, with a nonsignificant disadvantage for different-talker items among unpaired trials. This is interesting in that it suggests that listeners did not seem to use talker information strategically to gain an advantage in learning.

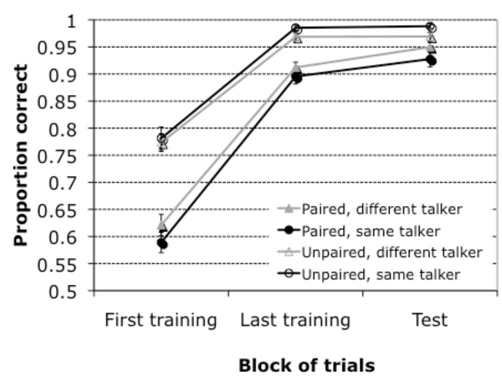


Figure 2. Accuracy during training and test phases, Experiment 1. Error bars are standard errors.

**Gaze fixations.** The dependent variable was target advantage: looks to the correct picture minus looks to the incorrect picture. (In the interests of space, we will not discuss analysis of the phonemically dissimilar "unpaired" trials.) Bearing in mind that signal-driven eye movements take about 200 ms to plan and execute (Hallett, 1986), and that the divergence point of paired words was 362 ms (+200 = 562 ms), we decided to examine target advantage two time windows prior to 550 ms: 200-400 ms, and 400-550 ms. We thus conducted an ANOVA with Talker Match (same, different) and Window (200-400ms, 400-550ms) as within-participants factors. There was an effect of Window, with Target Advantage increasing in the later window ($F(1,31) = 9.02$, $p = .005$). There was also an effect of Talker Match, with greater target advantage scores for different-talker pairs ($F(1,31) = 7.45$, $p = .01$). Finally, a Window x Talker Match interaction ($F(1,31) = 17.71$, $p = .0002$) indicated that the divergence between same- and different-talker trials was not significant in the first time window, but was significant in the second time window ($p = .0002$). In fact, target advantage for same-talker trials did not exceed 0. This implies that talker-specific information was used to predict word identity early on different-talker trials, before the words' phonemic point of divergence.

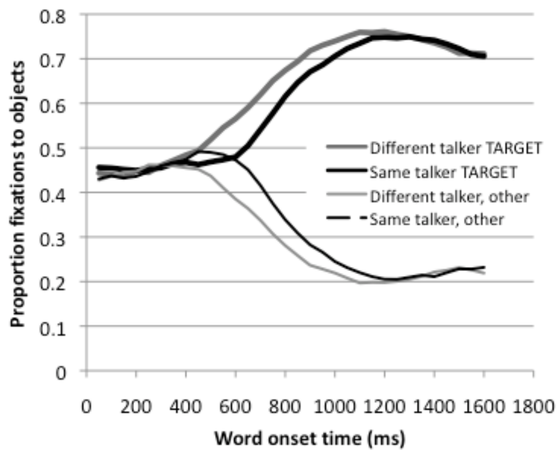


Figure 3. Gaze fixations to correct pictures (thick lines) and incorrect pictures (thin lines) on paired trials, Exp. 1.

## Discussion

Experiment 1 demonstrates both rapid and implicit use of talker information. Talker information is rapid in that it facilitates discrimination of words prior to the phonemic point of disambiguation, with about a 200 ms lead for talker-different pairs over talker-same pairs. Talker information was used implicitly in that it did not seem to improve accuracy—it merely facilitated processing. This null effect of accuracy contrasts with Experiment 2 of Creel et al. (2008), where talker information did improve listeners' accuracy levels. The likely reason for this difference is that the current vocabulary was easier to learn due to its greater resemblance to English words, and that listeners did not need to exhaust other cues (such as talker-varying acoustic attributes) in an attempt to learn the vocabulary. More generally, these accuracy and gaze fixation results replicate Creel et al. and support the idea that listeners necessarily store and utilize talker information during recognition of word-forms, consistent with models of memory suggesting that listeners' memories of word forms are highly acoustically accurate (e.g. Goldinger, 1998) rather than filtering for specific pieces of information.

However, a normalization account suggests that this use of talker information is a somewhat isolated phenomenon that occurs only when listeners do not have any other information at hand. Because words were learned in isolation, listeners likely had little time to acquire sufficient talker information to normalize upcoming input. Thus, in the next experiment, we test whether listeners continue to encode talker information as an integral part of a word's sound representation when words were presented more naturalistically in a carrier phrase. Here, listeners would have opportunity to calculate (and thus remove) talker-specific attributes from the words.

# Experiment 2

To assess whether learners still encode acoustically specific word representations even when they have sufficient information to account for talker-specific characteristics, we trained participants in Experiment 2 with words presented in sentence contexts ("Click on the X"). This provided sufficient acoustic context to calculate talker-specific attributes, which could potentially be used to normalize upcoming material. At test, the sentence context was removed and words were presented in isolation, as in Experiment 1. If learners encoded normalized forms at training, they should show attenuated or absent effects of talker variability at test. If, instead, they encoded talker information as an integral part of the signal despite the preceding context, then talker variability effects should remain strong.

## Method

**Participants.** N=32 participants from the same pool as in Experiment 1 took part.

**Stimuli.** The stimuli used were the same words as in Experiment 1, but were re-recorded in the frame sentence "Click on the X". The same two talkers from Experiment 1 recorded these sentences. The average length of the context was 511 ms. Data from a control experiment suggested that listeners were able to distinguish the two talkers within about the first 350 ms of the carrier phrase. The average divergence point of paired words (from word onset) was 390 ms.

**Procedure and equipment.** These matched Experiment 1, with the exception that participants were trained on full-sentence versions and tested on isolated words.

## Results

**Accuracy.** Participants took between two and five blocks of training ($m = 2.9$) to achieve 90% correct performance. Performance (Figure 4) improved over training trials, much as in Experiment 1. We again performed an ANOVA with Competitor Type (paired, unpaired), Talker Match (same talker, different talkers) and Block (first training block, last training block, test) as within-participants factors. Errors declined over Block ($F(2,62) = 302.32$, $p < .0001$). There were more errors on paired trials than unpaired trials ($F(1,31) = 242.96$, $p < .0001$). , A Block x Competitor Type interaction ($F(2,62) = 62.71$, $p < .0001$) suggested that paired trial accuracy came closer to unpaired trial accuracy in later blocks. There was an interaction of Talker Match x Competitor Type ($F(1,31) = 8.13$, $p = .007$), suggestive of a slight (nonsignificant) advantage for different-talker paired trials but a significant ($p = .04$) disadvantage for different-talker unpaired trials. Thus, there is not a clear pattern of an advantage for different-talker words.

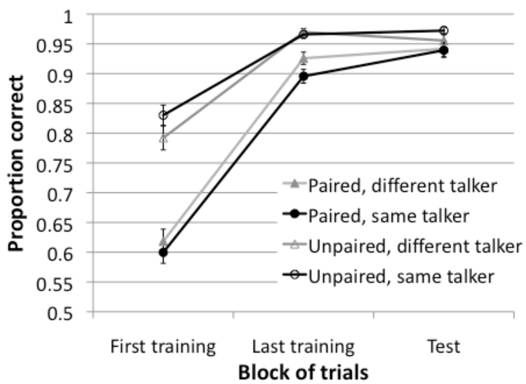


Figure 4. Accuracy on training trials and test, Experiment 2. Error bars are standard errors.

**Gaze fixations.** Gaze fixation patterns on unreinforced test trials (Figure 5) were quite similar to Experiment 1. We again analyzed two time windows in our ANOVA, 200-400 ms and 400-600 ms. The second window was extended to 600 ms because these words were slightly longer than those in Experiment 1 (591 ms). There was an effect of Window, with greater target-advantage scores in the later window ($F(1,31) = 5.3$, $p = .03$). There was also an effect of Talker Match, with greater target advantages for different-talker trials ($F(1,31) = 5.77$, $p = .02$). These factors did not interact ($F(1,31)=2.66$, $p = .11$). Compared to Experiment 1, the two Talker Match conditions did not differ in the 200-400 ms window, but differed in the 400-600 ms window ($p = .016$). These results confirm that listeners are storing talker-specific acoustic attributes about these words. This is particularly interesting in that listeners do not seem to have "factored out" talker identity during learning, even though they had ample spoken context over which to calculate talker identity prior to word onset.

## Discussion

In Experiment 2, we trained listeners to recognize talker-specific words in carrier phrases. This gave listeners the opportunity to extract talker-specific acoustic variables from the utterance prior to word onset. If they did this during training, we reasoned, then they might not encode talker-specific properties along with the word, meaning that they should not demonstrate talker-specific effects on the words alone at test. However, they did demonstrate talker-specific effects at test, with talker-different words being disambiguated sooner than talker-same words. This suggests that listeners did not normalize for talker information with more spoken context during learning. This strengthens the case that listeners routinely store analog representations of new words that do not subtract contextual acoustic variation.

# General Discussion

When listening to speech from a particular talker, the learner stores the acoustic variation that characterizes that talker's speech. Importantly, it seems that talker information is stored regardless of whether or not it occurs in context, even in cases where listeners should be able to extract "irrelevant" talker variability.

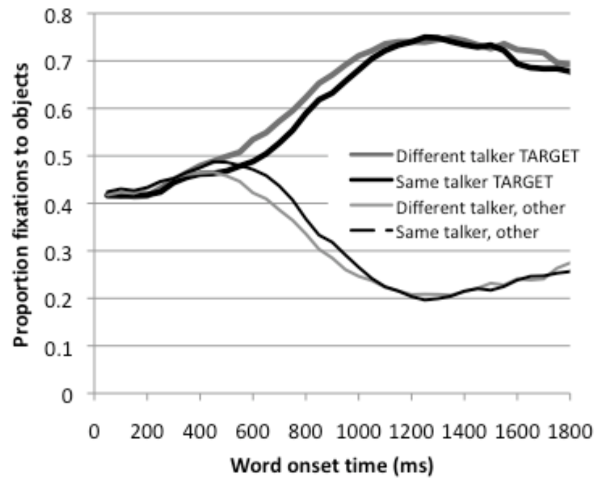


Figure 5. Fixations to correct (thick lines) and incorrect (thin lines) pictures, Experiment 2.

This research suggests not only that listeners can use nonphonemic variability to recognize words, but that listeners routinely store detailed acoustic information about words they hear. Such storage would allow the listener to retain a number of types of word form variability that might be useful in interpreting language in various contexts. For example, the learner could store accent-specific variability for various word forms. This is not equivalent to claims that the learner adjusts representations of word forms by accent (Dahan, Drucker, & Scarborough, 2008). The listener needs neither to adjust perception or representation, because the representations themselves include patterned variability.

These results have interesting implications with respect to normalization accounts. In essence, normalization is not needed: listeners need not to remove variability from the signal, but to predict the likely acoustic form of upcoming material. Thus, prior results indicating difficulty when talker identity changes rapidly (e.g. Magnuson & Nusbaum, 2007; Martin, Mullennix, Pisoni, & Summers, 1989) may reflect errors in prediction rather than the difficulty of repeated normalizations. That is, if the listener implicitly expects continuation in the same voice, an unexpected new voice will mismatch the prediction. Of course, this stands in contrast to data from Strange and colleagues (Jenkins, Strange, & Miranda, 1994) suggesting that vowel recognition is unimpaired when the talker is switched mid-syllable under a noise mask. Why rapid changes in talker cause processing difficulty in some cases and not in others is a topic for future research. It may be that while processing is equally accurate in some cases, processing ease suffers. The current data speak to this issue: while listeners were equally good at recognizing the words in different-talker and same-talker pairs, they were more rapid at recognizing the words in different-talker pairs. Future research will address this interesting pattern of data.

We have noted that child learners and second-language (L2) learners may be more sensitive than adult native speakers to talker variability. The current results indicate that even native-speaking adults are quite sensitive to talker variability. Thus, less expert word learners (children, for instance) may be even more profoundly affected by talker variation, as they do not have as much experience with the full range of talker variability. In ongoing research, we are exploring effects of talker specificity and talker diversity at earlier points in development.

In sum, listeners encode talker-relevant acoustic variation without removing context. Listeners do not store a "relative" form of a word that removes nonphonemic variability, but instead, a collection of acoustically-specific traces. This study extends previous work on talker specificity in word learning to more natural fluent-speech contexts.

## References


Allen, J. S., Miller, J. L., & DeSteno, D. (2003). Individual talker differences in voice-onset time. *Journal of the Acoustical Society of America, 119,* 544 – 552.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1996). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech perception. *Journal of the Acoustical Society of America, 101,* 2299 – 2310.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10,* 433 – 436.

Cornelissen, F.W., Peters. E., & Palmer, J. (2002). The Eyelink Toolbox: Eye tracking with MATLAB and the Psychophysics Toolbox. *Behavior Research Methods, Instruments & Computers, 34,* 613-617.

Creel, S. C., Aslin, R. N., & Tanenhaus, M. K. (2008). Heading the voice of experience: The role of talker variation in lexical access. *Cognition, 108,* 633 – 664.

Dahan, D., Drucker, S. J., & Scarborough, R. A. (2008). Talker adaptation in speech perception: Adjusting the signal or the representations? Cognition, 108, 710-718.

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

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.* New York: Wiley.

Houston, D. M., & Jusczyk, P. W. (2000). The role of talker-specific information in word segmentation by infants. *Journal of Experimental Psychology: Human Perception and Performance, 26(5),* 1570 – 1582.

Jenkins, J. J., Strange, W., & Miranda, S. (1994). Vowel identification in mixed-speaker silent-center syllables. *Journal of the Acoustical Society of America, 95,* 1030-1043.

Kraljic, T., & Samuel, A. G. (2007). Perceptual adjustments to multiple speakers. *Journal of Memory and Language, 56,* 1 – 15.

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.

Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y., & Yamada, T. (1994). Training Japanese listeners to identify English /r/ and /l/. III. Long-term retention of new phonetic categories. *Journal of the Acoustical Society of America, 96,* 2076 – 2087.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1990). Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America, 89,* 874 – 886.

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.

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

Martin, C. S., Mullennix, J. W., Pisoni, D. B., & Summers, W. V. (1989). Effects of talker variability on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 15,* 676-684.

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.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10,* 437 – 442.

Van Lancker, D., Kreiman, J., & Emmorey, K. (1985). Familiar voice recognition: Patterns and parameters. Part I. Recognition of backward voices. *Journal of Phonetics, 13*, 19-38.

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

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior & Development, 7,* 49 – 63.