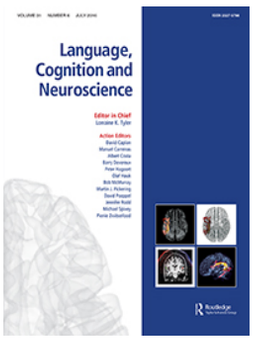

Routledge
Taylor & Francis Group

Language, Cognition and Neuroscience

ISSN: 2327-3798 (Print) 2327-3801 (Online) Journal homepage: http://www.tandfonline.com/loi/plcp21

Lexical access across talkers

Rochelle S. Newman

To cite this article: Rochelle S. Newman (2016) Lexical access across talkers, Language, Cognition and Neuroscience, 31:6, 709-727, DOI: 10.1080/23273798.2015.1136745

To link to this article: http://dx.doi.org/10.1080/23273798.2015.1136745

LANGUAGE, COGNITION AND NEUROSCIENCE, 2016
VOL. 31, NO. 6, 709–727
http://dx.doi.org/10.1080/23273798.2015.1136745

Routledge
Taylor & Francis Group

Lexical access across talkers

Rochelle S. Newman[a,b,c]

[a]Department of Hearing and Speech Sciences, University of Maryland, College Park, MD, USA; [b]Program in Neuroscience and Cognitive Science, University of Maryland, College Park, MD, USA; [c]Maryland Language Science Center, University of Maryland, College Park, MD, USA

ABSTRACT

The current paper examined whether lexical access might (ever) cross talker boundaries. In four cross-modal priming experiments, listeners made visual lexical decisions after hearing an auditory word or word pair. On some trials, the auditory signal was produced by a single talker; on other trials, a talker gender change occurred in the middle of the auditory sequence. Participants demonstrated priming for items crossing both word boundaries and talker changes: after hearing a male talker say my and a female talker say great, participants showed a speeded response to "geese" (related to "migrate"). Priming was based on the combination of syllables across talkers, not driven simply by the first word. Findings suggest that acoustic cues indicating multiple talkers are insufficient to disrupt lexical access, and activation of lexical representations is not limited to those occurring within a single-talker stream. This provides support for models that do not involve explicit pre-lexical segmentation.

ARTICLE HISTORY
Received 6 January 2015
Accepted 18 December 2015

KEYWORDS
Lexical access; lexicon; word boundary; segmentation; indexical

Introduction

A long-standing issue in the field of speech perception is how listeners identify the individual words within a fluent speech stream. Unlike written language, spoken language does not have an analog to "spaces" between words: although there are acoustic cues that can signal word boundaries, such cues are inconsistently available. Take, for example, a sentence such as *She saw two cans on the beach*. This could be interpreted as containing "two cans" or "toucans"; the phonetic sequence is the same in both cases. While there are sub-phonemic acoustic cues that can distinguish the two alternatives, these cues are probabilistic in nature. As a result, no one cue can be depended upon to distinguish whether in fact a word boundary has occurred. This can lead to at least momentary ambiguity in the interpretation of a sentence.

Models of spoken language recognition vary in how they address this potential word boundary ambiguity. Some models posit an explicit prelexical process of segmentation, in which sub-phonemic acoustic cues that frequently signal boundaries between words are used to identify likely words prior to lexical access. While such cues may not always be present in natural speech, and these prelexical segmentation hypotheses can be revised if subsequent lexical access fails, these models suggest that acoustic cues to word boundaries can reduce the number of potential interpretations of the signal that are initially activated in memory.

Other models suggest that segmentation is essentially a by-product of lexical competition among different potential parses (see Davis, Marslen-Wilson, & Gaskell, 2002 for discussion). Multiple items consistent with the speech input are initially activated in memory, and decided among through a process of constraint-satisfaction (see McQueen, 2005 for a review). These models suggest that many items will be initially activated, but that items that mismatch the signal will be quickly eliminated from consideration. Such models differ in how they weight different properties of the signal (e.g. acoustic, phonetic, lexical, etc.), as well as in what limits they place on this multiple activation. As a well-known example, the initial versions of the Cohort theory (Marslen-Wilson & Welsh, 1978) suggested that recognition occurs sequentially, and form-based lexical representations consistent with the stimulus input are activated only at word onset (or at the offset of the prior word). In contrast, other models, such as Trace (McClelland & Elman, 1986), Shortlist (Norris, 1994), and PARSYN (Luce, Goldinger, Auer, & Vitevitch, 2000) suggest that representations consistent with the input at any point during the lexical process become activated.

One critical source of evidence in these debates has been the results of studies investigating how listeners cope with word boundary uncertainty. Such studies often present a listener with a potentially ambiguous sequence, and then probe for activation of different interpretations. When activation is present for multiple

CONTACT Rochelle S. Newman rnewman1@umd.edu

parsing interpretations, it is taken as evidence in favour of models relying on lexical competition over models based on explicit segmentation (Davis et al., 2002); this is particularly the case when there are clear acoustic cues that are present in the signal and which presumably *could* have been used as a parsing cue. As an example, Gow and Gordon (1995) found that multiple-word sequences such as *two lips* activate concepts such as "tulips", despite the presence of acoustic cues to a word boundary in the middle of the sequence. They then suggest that their results support lexically based accounts of segmentation.

However, this argument depends on the clarity of those acoustic cues; models of word recognition based on prelexical segmentation may still predict activation of multiple possibilities when acoustic cues to segmentation are unclear. But at least when there are clear acoustic cues available, the presumption is that potential parsings that are ruled out by such acoustic cues should not lead to lexical activation.

The current studies examine one very obvious acoustic cue: that of a change in talker, and particularly a change in talker gender. Imagine a situation akin to Gow and Gordon but in which a male voice said the word *two* and a female talker subsequently said the word *lips* – in this situation, there are likely to be far greater acoustic differences across the two syllables than there would be when both words are spoken by the same talker. These larger acoustic differences may make it more clear that the two words should not combine to produce activation of the larger lexical representation, "tulips". If, however, activation of this longer sequence occurs, it would be strong evidence that acoustic cues to word boundaries are not sufficient to disrupt the process of lexical access (and thus that it is unlikely that segmentation based on such acoustic cues is occurring prior to lexical activation itself).

Before exploring this issue experimentally, we briefly summarise prior work on lexical access and segmentation within a single-talker situation, and then discuss how this might be extended to multiple-talker scenarios, and what implications such an extension might have for current theoretical debates.

Parsing the spoken signal

When listening to a person speaking, our percept is that we are hearing a series of individual words, presented in sequence like beads on a string. Yet an examination of the acoustic signal quickly shows that what hits the listener's eardrum is an ever-varying pattern of changes in air pressure, without any clear breaks indicating where one word ends and another begins (Cole & Jakimik, 1980; Klatt, 1980; Reddy, 1976).

This is particularly problematic because many long English words contain shorter words embedded within them (Luce, 1986; McQueen, Cutler, Briscoe, & Norris, 1985), as in the example *toucans* given earlier. Although there are acoustic juncture cues that might differentiate the two possible parsings (allophonic differences between word-initial /k/ and word-medial /k/, vowel durational differences, etc.; see, for example, Lehiste, 1960; Nakatani & Dukes, 1977; Quené, 1992), these cues are often quite subtle (Kim, Stephens, & Pitt, 2012), and their presence may in fact vary across different talkers or situations. As a result, models differ in how they predict such cues are used, including both debates on whether they are used to parse the signal prelexically vs. used as constraints during the process of lexical competition, and debates on the relative weighting of different types of cues (Mattys, White, & Melhorn, 2005).

A number of studies have explored the effect such cues have on lexical access, either by exploring activation of potential embedded words (e.g. "bone" in *trombone*; for some examples, see van Alphen & van Berkum, 2010; Bowers, Davis, Mattys, Damian, & Hanley, 2009; Luce & Cluff, 1998; Luce & Lyons, 1999; Norris, Cutler, McQueen, & Butterfield, 2006; Shillcock, 1990; Vroomen & de Gelder, 1997), or activation of longer sequences that cross apparent word boundaries (Christophe, Gout, Peperkamp, & Morgan, 2003; Gow & Gordon, 1995; Tabossi, Burani, & Scott, 1995). In general, such studies have found that even when acoustic cues to word boundaries are present, they do not preclude lexical access across them. As a result, spoken language often involves temporary ambiguities regarding the division of the fluent stream into individual words.

But such research has primarily taken place in situations in which a single individual is talking at one time. Here, the potential ambiguity revolves around whether a particular acoustic–phonetic sequence constitutes (all or part of) one word, or consists of the end of one word followed by the start of another (e.g. *toucans* vs. *two/cans*). Yet in many real-world situations, listeners may be trying to attend to one individual while other voices are speaking in the background, which adds to the complexity of the situation.

Although identifying individual talkers is itself a complicated process, listeners are quite good at distinguishing voices, and can even do so when the samples come from different languages (Wester, 2012). A change in talkers is likely to signal a concurrent boundary between words, in that two different talkers are presumably unlikely to collaborate in lexical production. Thus, it might make sense to posit a word boundary (or to halt lexical activation) wherever a change in talker occurs.

On the other hand, if segmentation is entirely the result of competition among multiple possible interpretations

of a signal, lexical activation might occur on the basis of whatever sequence of segments had been processed, regardless of the talker who produced them. Indeed, researchers have found equivalent identity priming for words across talkers as within talkers, at least when processing is fast (McLennan & Luce, 2005). After hearing the word *bacon* produced by one talker, listeners show speeded responses to the same word both when said again by the same talker and when said the second time by a novel talker, suggesting an abstract lexical representation. The authors suggested that, "indexical information in speech takes time to influence spoken word processing" (p. 316); if so, we might expect that activation of a word that crosses over talker streams would be feasible (although perhaps relatively uncommon in most real-world situations; see the final discussion for more on this point).

Thus, to summarise, a number of studies have suggested that acoustic cues to word boundaries alone may not be sufficient to disrupt the process of lexical access (for a recent study, see Kim et al., 2012). In particular, Gow and Gordon (1995) found activation for multisyllabic words such as "tulips" when listeners heard a speaker say the word *two* followed by *lips*. They used this finding to argue against an explicit process of prelexical segmentation. The current study uses a similar methodological approach, exploring whether the more obvious acoustic cues to talker identity can disrupt the process of lexical access. Different talkers are likely to be distinct in many ways: they will have different fundamental frequencies and formant structures, they may differ in dialect and/or speaking rate, their voices may differ in spatial location and in gender. In contrast to the subtle cues of word boundaries, talker change cues are likely to be very obvious acoustically – suggesting it would be less likely for lexical access to continue across the much more obvious boundaries induced by a change in talker. However, if lexical access takes place based on whatever legal phonetic signal hits the ears, it is conceivable that activation could continue to take place across talkers; such a finding would support theories of word recognition that place less emphasis on explicit prelexical segmentation. This would also have implications for episodic models of word recognition, which would presumably predict a greater mismatch between the input and stored representations for cross-voice than single-voice items.

Current research

The current paper examines whether lexical access might (ever) crosstalker boundaries. In a set of four experiments, we evaluate the extent to which lexical access is disrupted by acoustic boundary cues. These experiments use a cross-modal priming methodology, in which we examine the extent to which the auditory signal speeds processing in a subsequent lexical decision task. Participants will see a word or a nonword on a computer screen, and need to decide whether the item is a real word in the language as quickly as possible. Prior to this, listeners hear an auditory word or word pair. On some trials, this auditory sequence is semantically related to the item on the screen (and thus its prior activation could speed lexical access). In the critical experimental conditions, this sequence is only related if lexical access occurs across a talker boundary. For example, a participant might hear a male voice saying *two*, followed by a female voice saying *can*, and then be asked to decide whether "bird" is a real word in the language. If the participant's response to bird is faster in this case than when it follows voices saying *job* and *shoe*, it would suggest that the listener activated the word "toucan" despite the fact that the component syllables occurred in different voices.

The first study examines the effect of a talker change within a word. Bisyllabic words such as *cactus* are presented to listeners as potential primes, with a talker change occurring in the middle of the word. We examine whether the (combined) word is activated (whether it primes the subsequent lexical decision response). The size of the priming effect is compared when the word is spoken entirely in a female voice, entirely in a male voice, or with a talker change in the middle (either going from male to female or female to male). We expect to find priming when the word is presented in a single voice (either male or female); the question is whether such priming will be reduced when a talker change occurs mid-word.

The second study uses the same methodology, but uses two-word pairs. It is thus a replication and extension of Gow and Gordon's (1995) work, albeit using two-word phrases rather than complete sentences. Listeners hear sequences such as *kid knee*, in which the two words are spoken either by the same talker, or by two different talkers (one male, one female). We examine whether activation occurs for the combined sequence, "kidney" (whether listeners demonstrate priming for the related word "beans"). Gow and Gordon found lexical activation for "kidney" when a single talker said *kid* followed by *knee* – we explore whether we would also find such activation when the words are spoken by two different talkers (providing a second, more explicit cue for the location of a boundary between the words).

Our third study is similar to the second, but uses two-word sequences that cannot be combined to create a word. Critically, the first words in these sequences are

identical to those in Experiment 2. Thus, rather than hear *kid knee* followed by the visual word "beans", as in Experiment 2, listeners in Experiment 3 hear *kid go* followed by "beans". If priming results from Experiment 2 are driven by the first syllable (essentially, a form of cohort effect, as in Marslen-Wilson, 1987), we should see similar priming effects in this study. If, in contrast, the presence of both words is needed to result in priming for the subsequent visual item, it would provide stronger support for the notion that lexical access occurs across word and talker boundaries.

Our final experiment replicates Experiment 2, but reduces the proportion of primed sequences in the study, so as to better avoid any potential for strategic effects on the part of the listener. Thus, this study serves as a stricter test of the effect talker change has on lexical access.

As a group, these experiments examine the extent to which lexical access occurs across acoustic cues to a talker boundary (as in Experiment 1), word boundary (the single-voice conditions in Experiments 2 and 4), or both talker and word boundary in combination (the multiple-voice conditions in Experiments 2 and 4). The present studies thus explore the limits of lexical access.

Experiment 1

This first experiment was designed to evaluate whether lexical access would ever occur across a change in talker. As such, stimuli were designed to make it as likely as possible for lexical access to occur. Both a male and a female speaker recorded a series of bisyllabic words (such as *camel*, *rooster*, *second*, *maple*, etc.). Some words were presented as originally recorded, and others were cross-spliced such that there was a talker change in the middle of the word. We refer to the latter case as the cross-voice condition. In some cases, these words were potential primes for a lexical decision task, and we explored the extent to which different word types demonstrated priming. The full set of words is shown in Appendix 1.

We used two voices differing in gender so as to minimise the possibility that listeners would fail to hear the talker change. Although listeners are quite good at distinguishing voices, such differences can be overlooked when they are not the focus of attention, as shown in studies on change deafness (Vitevitch, 2003). Vitevitch found that when asked explicitly whether two male voices in his study differed, listeners were highly accurate (>90%); however, when given a shadowing task, 40% of the participants failed to notice a change between talkers. However, this type of change deafness primarily occurs when there is a single switch between talkers whose voices, although discriminable, are at least somewhat similar. Using two voices that differ in gender makes it unlikely that a change would go unnoticed. In addition, voice changes occurred on a large proportion of trials, and the fact that there would be different voices was explicitly pointed out to participants as part of the task instructions. Thus, although we encouraged lexical access across talkers by use of single, cross-spliced words, we also highlighted the change in talker.

Method

Participants

Thirty-two members of the University of Maryland community (23 female, 9 male) participated in exchange for course credit. All were right-handed, native speakers of English with no reported history of a speech, hearing, or attention disorder; an additional four participants were tested but were excluded for not meeting these criteria. Data from two additional participants were removed from analysis because they failed to meet a pre-established accuracy criterion (both always used one button, regardless of item lexicality). This study included 3 counterbalancing conditions (described below); of the 32 participants, 11 were tested in 2 of the 3 counterbalancing conditions, and 10 were tested in the third.

Stimuli

A total of 240 test stimuli were presented to listeners. Two talkers (one male, one female) were recorded producing words in isolation. Both speakers were recorded using a Shure SM81 microphone, and stimuli were digitised at a 44.1 kHz sampling rate and 16 bits precision, amplified, and stored on computer disk.

The author identified locations for cross-splicing in each word, based on visual and auditory inspection of the waveforms. Cut points were made at zero-crossings to avoid obvious clicks or discontinuities; either the first half of the word in the male voice was combined with the second half of the word in the female voice, or vice versa. Amplitude levels were first normalised such that all words had the same peak amplitude; then the two recordings were adjusted such that the adjacent sections before and after the cut point were comparable on the basis of the visual waveform. This resulted in 3 versions of each of the 240 words: 1 version entirely in the male voice, 1 entirely in the female voice, and 1 that that was cross-spliced. The cut points for all items are identified in the appendix; in some cases, one or both of the two portions happened to make a word by itself (*army* = *are, me*), but this was not consistent (*summer* = *suh, mer*). It is worth pointing out that only the cross-voice

items were cross-spliced; although this does lead to a potential confound, it makes it even less likely that we would find similar results in the cross-voice items compared to the male or female items.

Procedure

Participants took part individually in a lexical decision task using Psyscope (Cohen, MacWhinney, Flatt, & Provost, 1993). On each trial, participants first saw a fixation cross for 500 ms. They then heard a two-syllable auditory word over headphones. A visual item appeared on the screen 50 ms after the offset of the auditory word; participants indicated whether this visual item was a real word in English or a nonsense word by pressing an appropriately labelled button on a computer-controlled response box. Participants were asked to respond as quickly and accurately as possible to these items, and responded *word* with their preferred, right hand. Stimulus presentation and response collection were controlled by computer. Following the participant's response, there was a 1000 ms inter-trial interval before the onset of the next trial.

Auditory stimuli were presented at a comfortable listening level over circumaural headphones. Visual items were presented in normal, lower-case typeface (Chicago, 24 point) on a high quality monitor. Prior to the experimental set, participants were given a block of 18 practice trials using similar, but additional, items (6 male, 6 female, 6 cross-voice) to familiarise them with the task.

The test phase consisted of 240 trials; half (120) of the visual items were real words, and half were nonwords. Nonwords were orthographically similar to real words (e.g. vocket–rocket; chaggy–shaggy; tobe–robe; elg–elk; jile–mile, etc.). For the real words, half (60) were preceded by an auditory sequence that was potentially related (a prime), and half were preceded by an unrelated auditory word (thus, the word "movie" might be preceded by *popcorn* (a related word) or by *pickle* (an unrelated word)). All related words were semantic associates selected on the basis of pilot testing. All nonwords were preceded by a similar auditory sequence (that is, there was nothing fundamentally different between the auditory sequences that preceded visual words (e.g. *tulip*, *dolphin*, *heavy*, …) and those that preceded visual nonwords (e.g. *baseball*, *captain*, *cricket*, …)).

The auditory sequence was entirely in a female voice one-third of the trials (20 of which were followed by a related word (primed trials), 20 followed by an unrelated word (unprimed trials), and 40 followed by a nonword (foil trials)), entirely in a male voice on one-third, and cross-voice on one-third (with half being male voice first, and half female voice first); these ratios were identical for primed, unprimed, and nonword trials. No participant heard the same auditory prime twice, but, across participants, each prime occurred in all three conditions (male, female, and cross-voice). In order to maintain this counterbalancing, participants were randomly assigned to one of three different sets of stimuli; order of the stimuli was then randomised for each participant.

Stimuli were rotated through conditions both within and across participants: thus, the auditory word *popcorn* occurred in a female voice for one-third of the participants, a male voice for one-third of the participants, and a cross-voice condition for one-third of the participants. No participant heard the same auditory word more than once. The visual item "movie" appeared twice for each participant: once primed (preceded by *popcorn*) and once unprimed (preceded by an unrelated word).

Reaction times (RTs) and accuracy were calculated for each item for each participant, measured from the onset of the visual item to the button press. Any response time that was greater than 2000 ms was removed, as was any response time less than 50 ms (total of 28 out of 7680 trials, or <0.4%). Because of a stimulus error, one intended trial was not given (the unprimed version to the word "sad"); this item was removed from the items analysis, only. Finally, any RT greater than 2 standard deviations from the participant's mean was removed from the subjects analysis (accounting for 103 of the target trials, or 5.4%). Analyses were done both by subjects, and by items.

Results and discussion

Average lexical decision accuracy was 97.8%, demonstrating that participants were extremely accurate at the task, and at near-ceiling performance.

We conducted a 2 (prime relatedness: related or unrelated) × 3 prime voice (male only, female only, or cross-voice) analysis of variance (ANOVA) on participants' RT data. Overall, we found a significant effect of prime relatedness (both $p < .0001$; see both Table 1 for statistical analyses and Figure 1): participants responded in 556 ms for related items, but 594 ms for unrelated (unprimed) items. There was an overall effect of prime voice in the subjects analysis only ($p_1 < .006$; $p_2 = .06$), such that participants were somewhat faster at responding to the visual item when the preceding voice was female than when the preceding voice was male, with response times falling intermediate between the other two conditions when the preceding item was cross-voice (male voice: average RT = 585 ms; female voice = 564 ms; cross-voice = 576 ms). The effect size suggests this is not a very meaningful difference. It is not clear

Table 1. Statistical results from Experiment 1.

Effect	By subjects	By items
RT data		
Prime relatedness	$F_1(1,31) = 30.31$, $p < .0001$, $\eta_p^2 = .494$	$F_2(1,58) = 35.90$, $p < .0001$, $\eta_p^2 = .382$
Prime voice	$F_1(2,62) = 5.71$, $p < .006$, $\eta_p^2 = .156$	$F_2(2,116) = 2.85$, $p = .062$, $\eta_p^2 = .047$
Interaction	$F_1(2,62) = 2.20$, $p = .12$, $\eta_p^2 = .066$	$F_2(2,116) = 1.01$, $p = .37$, $\eta_p^2 = .017$
Relatedness effect after a male prime	$t_1(31) = 5.56$, $p < .0001$	$t_2(59) = 4.86$, $p < .0001$
Relatedness effect after a female prime	$t_1(31) = 3.26$, $p < .005$	$t_2(59) = 3.59$, $p < .0001$
Relatedness effect after a cross-voice prime	$t_1(31) = 2.74$, $p < .02$	$t_2(58) = 3.09$, $p < .005$
Accuracy data		
Prime relatedness	$F_1(1,31) = 2.56$, $p = .12$, $\eta_p^2 = .076$	$F_2(1,58) = 0.70$, $p = .41$, $\eta_p^2 = .012$
Prime voice	$F_1(2,62) = 4.64$, $p = .013$, $\eta_p^2 = .13$	$F_2(2,116) = 2.31$, $p = .104$, $\eta_p^2 = .038$
Interaction	$F_1(2,62) = 1.31$, $p = .28$, $\eta_p^2 = .046$	$F_2(2,116) = 1.99$, $p = .14$, $\eta_p^2 = .033$
Relatedness effect after a male prime	$t_1(31) = 0.85$, $p = .40$	$t_2(59) = 0.61$, $p = .55$
Relatedness effect after a female prime	$t_1(31) = 2.63$, $p = .013$	$t_2(59) = 1.93$, $p = .058$
Relatedness effect after a cross-voice prime	$t_1(31) = 0.16$, $p = .40$	$t_2(58) = 0.92$, $p = .36$

how to interpret this; the two talkers did not speak at the exact same rate, but this voice effect was inconsistent across items. Perhaps variation in talking rate slowed the participants' responding to trials in which the male voice spoke.

Most important, however, there was no interaction between prime voice and prime relatedness ($p_1 = .12$; $p_2 = .37$). Apparently, the size of the relatedness effect was relatively similar between conditions. Looking at the three conditions separately, we find a 54 ms relatedness effect following the male voice prime, compared with a 35 ms relatedness effect following the female voice prime, and a 28 ms relatedness effect following the cross-voice condition, all significant. While the size of the relatedness effect appears slightly different across conditions, this is primarily driven by a larger effect in the male voice; the cross-voice and female conditions are quite similar. Moreover, the lack of any interaction suggests that what differences there were across voices are inconsistent. Importantly, the relatedness effects are similar across both items and subjects, suggesting that they are not being driven by particular stimuli.

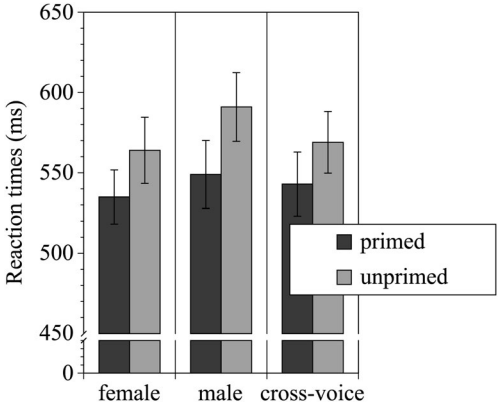

Figure 1. RTs from Experiment 1 to visual words preceded by related (primed) and unrelated (unprimed) words, spoken either by a single voice or a voice combination. Error bars represent standard error.

The accuracy data (using arcsin transforms) showed an effect of prime voice (greater accuracy for visual words following a female voice prime (98.8%) than following a male (97.9%) or cross-voice (97.5%) prime), but only in the subjects analysis ($p_1 = .013$; $p_2 = .10$). There was no overall effect of prime relatedness ($p_1 = .12$; $p_2 = .41$) and no interaction ($p_1 = .28$; $p_2 = .14$). Looking at the three conditions separately as planned comparisons, we found an effect of priming only in the female voice in the subjects analysis (99.5% vs. 98.1% accuracy). Neither the items entirely in the male voice (98.1% vs. 97.6%) nor the cross-voice items (97.5% for each) showed any accuracy differences. This may be the result of the near-ceiling effects on this measure.

To summarise, then, this first experiment demonstrated similar amounts of priming when the prime was spoken in a single voice as compared to when there was a talker change in the middle of the word. This effect was found in the RT data both when analysed by subjects and by items, and suggests that lexical access consistently occurred despite the talker change. This would be unexpected if there was an explicit segmentation process occurring prior to lexical access that was sensitive to acoustic discontinuities in the signal. It would also be unexpected if lexical access occurred on the basis of talker-specific episodic traces, since presumably participants' stored representations would better match single-voice items than cross-voice items.

However, there are several limitations in this study that limit its interpretation. First, the stimuli used were such that, in general, the sequence was only a legitimate word when heard as a whole unit. That is, for a word such as *Easter*, neither "ees" nor "tir" are themselves a word; the sequence was only legal if the two portions were combined. Parsing it as two units led to an illegal (or at least, nonlegitimate) lexical item. This could have induced unusual listening strategies in the participants that would not carry over to other environments. Indeed, one possible interpretation of the present results is that lexical access proceeds in such a way as to avoid stranding illegal nonword sequences, and that this led listeners to instead incorporate speech across voices. Prior research has generally found otherwise,

however; although listeners avoid parsing in ways that strand items that are *impossible* as a word (such as a single consonant; Norris, McQueen, Cutler, & Butterfield, 1997), they do not avoid parsing in ways that strand items that are simply not known as words or are unlikely to be words (Newman, Sawusch, & Wunnenberg, 2011), as was the case here.

A second limitation is that the words had been produced as a whole entity. For example, both speakers produced the word "apple" as a whole word before they were cross-spliced. This means that there were acoustic cues to continuity between the first and second syllables – the syllables were produced with coarticulation into and out of the other syllable. Thus, the items had a talker change, but did not simultaneously include cues to the presence of a word boundary. This may have encouraged listeners to treat the word as a (combined) whole.

Third, only the cross-voice items were cross-spliced. This might have resulted in some unnaturalness specific to those stimuli. This should have resulted in it being LESS likely to find priming in the cross-voice items, rather than more likely, and thus cannot explain the similar results across conditions. Nonetheless, it is not ideal methodologically. To address these concerns, Experiment 2 explores whether lexical access would also occur across two separate words spoken by different talkers.

Experiment 2

The first experiment demonstrated that lexical access would "cross-over" a change in talker. However, as noted above, there were several limitations in that study that make interpreting these results somewhat difficult. Experiment 2 addresses these concerns by examining whether such cross-voice lexical access would occur even when the speech signal in each voice was a legitimate word. Gow and Gordon (1995) demonstrated that when a single talker spoke the words *two* and *lips* (as part of a fluent sentence), listeners activated the larger word, "tulips". Experiment 2 explores whether this would still occur when the two words were spoken in different voices.

However, Gow and Gordon's study presented the words in the context of fluent speech. While full sentences may be more ecologically valid than are two-word pairs, they also involve coarticulation among adjacent syllables, even when those syllables are separate words. Moreover, some research (Norris et al., 2006) suggests that semantic (associative) priming is more difficult to find in a fluent context unless the task encourages listeners to attend specifically to that region of the sentence (such as by truncating the sentence early, as Gow and Gordon did). Since our goal was to compare the amount of semantic priming within a voice to that across voices, we chose to use auditory stimuli that more consistently result in priming effects (i.e. individual words). Moreover, by recording the words individually in list format (in random order), we also eliminate any coarticulation effects across the words. This should encourage segmentation of the two words (Mattys, 2004), thus reducing the likelihood of priming of the integrated sequence. Thus, if lexical access across voices depends on continuity of production (either based on formant coarticulation or prosodic patterning), then the effect should disappear in this study.

Thus, Experiment 2 addresses each of the limitations from Experiment 1: the items all contain acoustic cues to a word boundary (in addition to any cues for a talker change); all stimuli are made from concatenated words (so there is no difference in cross-splicing); and all of the component pieces are individual words (such that positing a word boundary does not strand any illegal nonword sequences). We examine whether lexical access occurs despite these multiple cues to the presence of a word boundary.

Method

Participants

Twenty-eight members (19 female, 8 male; 1 unreported) of the University of Maryland community took part in this experiment. All participants were right-handed, native speakers of English with no reported history of a speech, hearing, or reading disorder. Participants received course credit or a cash payment for their participation. The data from one additional participant was excluded for missing data; two others were excluded for experimenter error ($n = 1$) or for failing to complete the study. This left a total of seven participants in each of four counterbalancing conditions.

Stimuli

Stimuli were based on word combinations used by Gow and Gordon (1995), although additional word pairs were added to their set. See Appendix 2 for a complete list of word pairs. A total of 480 stimuli were presented to listeners. Two talkers (one male, one female, both unaware of the purposes of the study) were recorded producing these words in isolation. The words were presented to them in a pseudo-random order, such that the items that would be used as target stimuli were not adjacent in the list when recording. Pairs of these words were then concatenated, with 100 ms between words. The existence of a short gap aided in the perception that

these were individual words in a list, rather than a single sequence. This method of recording means that the words were produced as isolated words (in a list fashion); there was no coarticulation going from one word into the other, because they had not been produced adjacent to one another.

Procedure

The procedure used was similar to that in Experiment 1. On each trial, participants first saw a fixation cross for 500 ms. They then heard a two-word sequence played over headphones. A visual item appeared on the screen 50 ms after the offset of the auditory sequence. The participant indicated whether the visually presented item was a real word in English or a nonsense word by pressing an appropriately labelled button on a computer-controlled response box. Participants were asked to respond as quickly and accurately as possible to these items, and responded *word* with their preferred, right hand. RTs were measured from the onset of the visual item to the button press; RTs longer than 2000 ms or less than 50 ms were ignored (this accounted for 0.5% of the data). Following the participant's response, there was a 1000 ms inter-trial interval before the onset of the next trial.

There was a 12-trial practice block (consisting of 3 items in each voice condition, with a total of 6 words and 6 nonwords), followed by 480 test trials divided into 4 blocks of 120 trials each. Half of the visual items were real words, and half were not (foils). For the real words, half were preceded by an auditory sequence that was potentially related to it (a related prime), and half were preceded by an unrelated auditory sequence. Thus, the word "ship" might be preceded by the two-word sequence *cap-tin* (which could form a related word, captain) or by *quick-kite* (which could not create a related word).

The same auditory words that occurred in related sequences also occurred in unrelated sequences in pairings that could not be combined (thus, "cap-tin" could be heard as word *captain*, but *drag-tin* and *cap-miss* cannot be combined in this manner). As a result, only one-fourth of the two-word sequences could potentially be combined into a single, legal word. But the same auditory words were presented both as part of combinable sequences and as part of noncombinable sequences. The only potential for related-word priming in this study came from combinable sequences, not from individual words (e.g. while *car pet* is related to the visual word "rug", the word *car* was never presented with the word "auto"; the only potential for related meanings between the auditory and visual items came from those potentially combinable sequences).

The auditory word pairs were both spoken by a female voice on 25% of the trials (120 total) for each type (primed = 30, unprimed = 30, foils = 60), and both by a male voice on 25%. Of the other trials, half had a female word followed by a male, and half a male followed by a female. Thus the total set of individual words was spoken by a male voice 50% of the time, and the female voice 50% of the time. (This ratio of different trial types is slightly different than in Experiment 1; in Experiment 1, the cross-voice items made up one-third of the stimuli; here, there are no cross-voice words, but 50% of the trials contain two different voices, each saying one word.) No participant heard the same auditory two-word sequence twice; thus, there were four different conditions, and participants were assigned randomly to one of the four conditions.

Rotation of items occurred both within and across participants: each visual word occurred twice per participant, once preceded by a related item and once by an unrelated item. Each auditory word also occurred twice per participant, once as part of a legal combination (*win-dough* → window) and once as part of an uncombinable sequence (*four-dough*). And each critical combination occurred in each talker combination (male–male, female–female, male–female, and female–male) across participants.

Results and discussion

Average lexical decision accuracy was 96.99%, demonstrating that participants were extremely accurate at the task. To explore the effects of talker on priming, we conducted $2 \times 2 \times 2$ ANOVAs, with the variables of prime relatedness (related vs. unrelated), number of voices (both words in a single voice vs. two different voices), and first voice (female first or male first). The last variable distinguishes between male/male and female/female versions in the single-voice condition, and between male/female and female/male versions in the two-voice condition, but is not a variable in which we are particularly interested. The critical comparison is the interaction between priming and the number of voices.

Analyses were performed both by subjects and by items; analyses were done removing outliers from the subjects' RT data (186 of the target trials, or 2.77%) and using arcsin transforms for accuracy. RT data are shown in Figure 2, and statistical results in Table 2.

Overall, we found a significant effect of prime relatedness in the RT data (both $p < .0001$), with related items being responded to faster than unrelated items (542 vs. 562 ms). There were no other significant main effects or interactions. The critical interaction of

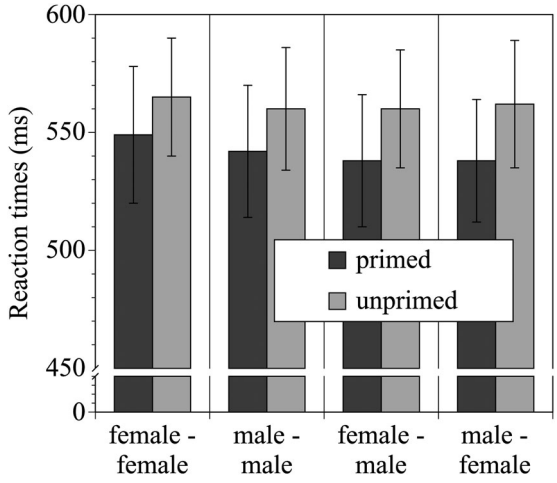

Figure 2. RTs from Experiment 2 to visual words preceded by word pairs that could either be combined into a related word (i.e. *two-lips* related to "flower"; primed) or which could not (unprimed), spoken either by a single voice or a voice combination. Error bars represent standard error.

relatedness by number of voices did not approach significance ($p_1 = .29$; $p_2 = .15$), indicating that the relatedness effect was no less for the cross-voice conditions.

Looking at the four conditions separately, we find a marginal 15-ms relatedness effect in the female voice ($t_1(27) = 1.97$, $p = .059$; $t_2(117) = 1.99$, $p = .049$), a significant 18-ms relatedness effect in the male voice ($t_1(27) = 3.04$, $p = .005$; $t_2(117) = 3.07$, $p = .003$), a 22-ms relatedness effect in the female–male cross-voice items ($t_1(27) = 4.12$, $p = .0003$; $t_2(117) = 3.48$, $p < .001$), and a 24-ms relatedness effect in the male–female cross-voice items ($t_1(27) = 4.11$, $p = .0003$; $t_2(117) = 4.02$, $p < .0005$). There is thus no indication that the relatedness effect was reduced for the cross-voice items.

Table 2. Statistical results from Experiment 2.

Effect	By subjects	By items
RT data		
Prime relatedness	$F_1(1,27) = 27.11$, $p < .0001$, $\eta_p^2 = .501$	$F_2(1,117) = 31.74$, $p < .0001$, $\eta_p^2 = .213$
Number of voices	$F_1(1,27) = 2.10$, $p = .16$, $\eta_p^2 = .072$	$F_2(1,117) = 0.36$, $p = .55$, $\eta_p^2 = .003$
First voice	$F_1(1,27) = 0.54$, $p = .47$, $\eta_p^2 = .02$	$F_2(1,117) = 0.01$, $p = .92$, $\eta_p^2 < .001$
Relatedness × number of voices	$F_1(1,27) = 1.17$, $p = .29$, $\eta_p^2 = .041$	$F_2(1,117) = 2.15$, $p = .146$, $\eta_p^2 = .018$
Relatedness × first voice	$F_1(1,27) = 0.16$, $p = .689$, $\eta_p^2 = .006$	$F_2(1,117) = 1.22$, $p = .27$, $\eta_p^2 = .010$
Number of voices × first voice	$F_1(1,27) = 1.17$, $p = .29$, $\eta_p^2 = .041$	$F_2(1,117) = 1.45$, $p = .23$, $\eta_p^2 = .012$
Three-way interaction	$F_1(1,27) = 0.01$, $p = .92$, $\eta_p^2 < .001$	$F_2(1,117) = 0.01$, $p = .94$, $\eta_p^2 < .001$
Accuracy data		
Prime relatedness	$F_1(1,27) = 7.94$, $p = .009$, $\eta_p^2 = .227$	$F_2(1,117) = 9.72$, $p = .002$, $\eta_p^2 = .077$
Number of voices	$F_1(1,27) = 0.66$, $p = .42$, $\eta_p^2 = .024$	$F_2(1,117) = 0.44$, $p = .51$, $\eta_p^2 = .004$
First voice	$F_1(1,27) = 2.47$, $p = .127$, $\eta_p^2 = .084$	$F_2(1,117) = 1.20$, $p = .275$, $\eta_p^2 = .010$
Relatedness × number of voices	$F_1(1,27) = 2.48$, $p = .13$, $\eta_p^2 = .084$	$F_2(1,117) = 1.31$, $p = .255$, $\eta_p^2 = .011$
Relatedness × first voice	$F_1(1,27) = 0.28$, $p = .60$, $\eta_p^2 = .01$	$F_2(1,117) = 0.24$, $p = .63$, $\eta_p^2 = .002$
Number of voices × first voice	$F_1(1,27) = 1.73$, $p = .20$, $\eta_p^2 = .06$	$F_2(1,117) = 0.01$, $p = .96$, $\eta_p^2 < .001$
Three-way interaction	$F_1(1,27) = 2.06$, $p = .16$, $\eta_p^2 = .07$	$F_2(1,117) = 1.16$, $p = .283$, $\eta_p^2 = .010$

Unlike in Experiment 1, the accuracy data here mirrored the RT data, with a main effect of prime relatedness ($p_1 = .009$; $p_2 = .002$), and no other main effects or interactions. Thus, all of the data showed a similar pattern: there was an effect of relatedness across the four conditions, and the change in voice led to no reduction in the amount of this advantage.

These results mirror those reported by Gow and Gordon (1995): when listeners hear a two-word sequence such as *tree owes*, they activate not only those words, but also activate longer words that contain those two syllables ("trios"). And perhaps surprisingly, they continue to do so even when the two words are spoken by two separate talkers of different genders.

There is one major methodological difference between the present study and that of Gow and Gordon, however: Gow and Gordon embedded their words in fluent sentences. They also truncated their sentences after the prime occurred. We chose not to use that approach based on results by Norris et al. (2006). Norris et al. suggest that in a fluent speech sequence, associative priming reflects conceptual activation associated with the utterance as a whole, rather than reflecting lexical access to the individual words. Thus, sentences may simply not show priming to individual words that are not closely tied to the utterance-level interpretation. In order to compare sequences such as *two lips* vs. *tulips* in sentence context, the sentences themselves must contain no other semantic cues to either interpretation, and such sentences (according to Norris et al.) might often fail to show priming, even when priming would occur for words in isolation. Moreover, Norris et al. suggest further that construction of an utterance-level interpretation will depend on the particular demands facing the listener at that moment; for example, Gow and Gordon's priming effects may have occurred because the truncation of the sentences part-way through focused attention strategically on the final word heard. Presenting a talker change in the middle of a sentence might be likely to also induce strategic effects of this sort, and this might differ across talkers. Since our goal in the present study was not to explore issues of conceptual processing *per se*, we opted to avoid sentential contexts, which might make interpretation more difficult. Rather, our intent was simply to find a situation in which, for single words produced by

an individual talker, priming consistently occurred, and then to compare the extent of this priming across different numbers of talkers. Moreover, by using isolated words, recorded in list format, we avoided the types of coarticulation across word boundaries that would be likely in a fluent speech context. Thus, we extend Gow and Gordon's findings to show that cross-boundary lexical access is not limited to situations in which there are coarticulatory cues that would encourage conjunction.

One concern, however, is that the effect may be driven by the first word alone. Marslen-Wilson (1987) and Marslen-Wilson and Welsh (1978) proposed a model of lexical access (the Cohort model) in which the onset of a speech sequence prompts activation of all words beginning with that sequence. Thus, hearing part of a word is sufficient to generate lexical access in this model. Regardless of the perceived accuracy of this model overall, it is nonetheless possible that the presence of the first word in the current study is sufficient to activate the larger word, without any activation occurring across a boundary. Indeed, Marslen-Wilson (1987; see also Zwitserlood, 1989) found that listeners showed activation for both "captain" and "captive" after hearing *cap*. Perhaps in the present study, hearing *kid* is sufficient to activate "kidney"; if so, the fact that *knee* was spoken in a different voice would not matter – activation would not truly be occurring across voices in this case.

To examine this, Experiment 3 mirrors Experiment 2 except that the second words in the various auditory sequences were shuffled. Thus, rather than hear *kid knee* and *car go*, participants might hear *kid go* and *car knee*. If the first syllable is sufficient to generate lexical access or maintain it, we should continue to see priming in this situation. If, however, the results from Experiment 2 are based on the legality of the combination of the two words, we would not expect to see priming with these mismatched combinations.

Experiment 3

In this experiment, we examine whether we continue to get priming for a visual item related to a longer sequence that begins the same way as do those in Experiment 2. If the first item in the two-word sequence is driving the effect seen in the prior experiment, we should find similar priming regardless of the identity of the second word.

Method

Participants

Thirty-two members of the University of Maryland community took part in this experiment. All participants were right-handed, native speakers of English with no reported history of a speech, hearing, or reading disorder. Participants received course credit or a cash payment for their participation. The data from an additional nine participants were not analysed for the following reasons: not being a native speaker ($n = 5$), being left-handed ($n = 2$), experimenter error ($n = 1$) or having ADD ($n = 1$).

Stimuli

Stimuli were identical to those in Experiment 2, except that the second words of the two-word sequences were shuffled, resulting in two-word sequences that could not make larger multisyllabic lexical items. On critical trials, the visual item was related to a larger would that could have been made from a continuation of the first auditory word. (For example, participants hear *ran muse*, and saw a visual item related to "random" – a word that could have been made from a continuation of the first auditory word.)

Procedure

The procedure used was identical to that in Experiment 2.

Results and discussion

Average lexical decision accuracy was 96.5%, demonstrating that participants were very accurate at the task. To explore the effects of talker on priming, we conducted $2 \times 2 \times 2$ ANOVAs, in the same manner as in Experiment 2, both by subjects and by items; analyses were done removing outliers from the RT data and using arcsin transforms for accuracy. RT data are shown in Figure 3 and statistical results in Table 3.

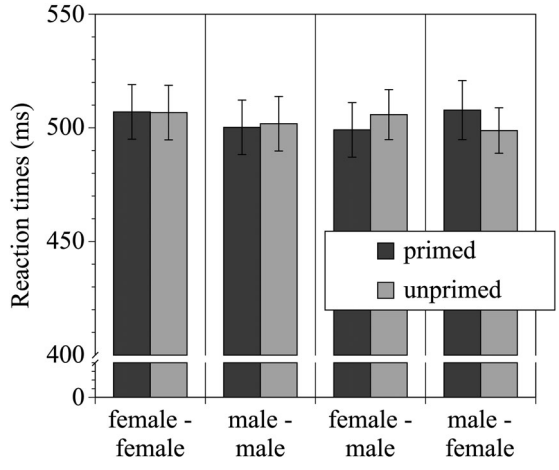

Figure 3. RTs from Experiment 3 to visual words preceded by word pairs containing the same initial words as in Experiment 2, spoken either by a single voice or a voice combination. Error bars represent standard error.

Table 3. Statistical results from Experiment 3.

Effect	By subjects	By items
RT data		
Prime relatedness	$F_1(1,31) = 0.03$, $p = .88$, $\eta_p^2 = .001$	$F_2(1,119) = 0.38$, $p = .54$, $\eta_p^2 = .003$
Number of voices	$F_1(1,31) = 0.28$, $p = .60$, $\eta_p^2 = .009$	$F_2(1,119) = 2.36$, $p = .13$, $\eta_p^2 = .019$
First voice	$F_1(1,31) = 2.03$, $p = .16$, $\eta_p^2 = .06$	$F_2(1,119) = 0.31$, $p = .58$, $\eta_p^2 = .003$
Relatedness × number of voices	$F_1(1,31) = 3.22$, $p = .08$, $\eta_p^2 = .094$	$F_2(1,119) = 1.20$, $p = .28$, $\eta_p^2 = .01$
Relatedness × first voice	$F_1(1,31) = 0.15$, $p = .70$, $\eta_p^2 = .005$	$F_2(1,119) = 0.15$, $p = .70$, $\eta_p^2 = .001$
Number of voices × first voice	$F_1(1,31) = 3.98$, $p = .06$, $\eta_p^2 = .11$	$F_2(1,119) = 1.17$, $p = .28$, $\eta_p^2 = .01$
Three-way interaction	$F_1(1,31) = 4.83$, $p = .036$, $\eta_p^2 = .135$	$F_2(1,119) = 0.17$, $p = .68$, $\eta_p^2 = .001$
Accuracy data		
Prime relatedness	$F_1(1,31) = 1.02$, $p = .32$, $\eta_p^2 = .032$	$F_2(1,119) = 0.05$, $p = .83$, $\eta_p^2 < .001$
Number of voices	$F_1(1,31) = 0.05$, $p = .82$, $\eta_p^2 = .002$	$F_2(1,119) = 0.20$, $p = .65$, $\eta_p^2 = .002$
First voice	$F_1(1,31) = 2.20$, $p = .15$, $\eta_p^2 = .066$	$F_2(1,119) = 1.01$, $p = .32$, $\eta_p^2 = .008$
Relatedness × number of voices	$F_1(1,31) = 0.55$, $p = .46$, $\eta_p^2 = .017$	$F_2(1,119) = 0.64$, $p = .43$, $\eta_p^2 = .005$
Relatedness × first voice	$F_1(1,31) = 3.53$, $p = .07$, $\eta_p^2 = .102$	$F_2(1,119) = 1.41$, $p = .24$, $\eta_p^2 = .012$
Number of voices × first voice	$F_1(1,31) = 0.10$, $p = .76$, $\eta_p^2 = .003$	$F_2(1,119) = 0.40$, $p = .53$, $\eta_p^2 = .003$
Three-way interaction	$F_1(1,31) = 0.55$, $p = .47$, $\eta_p^2 = .017$	$F_2(1,119) = 0.00$, $p = .99$, $\eta_p^2 < .001$

Overall, we found no effect of prime relatedness in the RT data; this is the first analysis that did not show a relatedness effect, and it is worth noting that the effect was not even approaching significance ($p_1 = .88$; $p_2 = .54$). There were also no other significant main effects or interactions, except for the three-way interaction between priming, number of voices, and first voice, which was significant by subjects only. Looking at this three-way interaction more closely, it appears to be the result of the fact that the female–male cross-voice items showed a 6 ms advantage for related visual items, while the male–female cross-voice items showed a 9 ms disadvantage for related items. The male and female voice items showed no differences (1.6 ms difference in the male voice, −0.3 ms difference in the female voice). Although the three-way interaction is significant, it does not appear to demonstrate any true effect of relatedness, and the differences found are substantially smaller than the consistent relatedness advantages found in Experiment 2. Accuracy data likewise showed no main effects nor any interactions; the critical interaction did show a marginal effect in the subjects analysis, but this trend was in the opposite direction of that expected, with poorer accuracy in the primed conditions for the cross-voice items (on the order of 0.75% difference), and no differences in the single-voice items (0.1% and 0.4% differences).

Thus, the current results show no evidence of any prime relatedness effects in this experiment, suggesting that hearing the first word alone is not sufficient to generate a speeded response to the visual item. This, in turn, implies that the effects found in Experiment 2 were likewise not being driven by the first word alone. It is possible that there had been some initial priming of the longer potential sequence that was eliminated after the mismatching second word occurred; this would be in line with the predictions of the cohort model (Marslen-Wilson & Welsh, 1978). But clearly the results are substantially different when a listener hears "car" in a female voice followed by "knee" in a male voice in this experiment, than when hearing "car" in a female voice followed by "go" in a male voice in the prior experiment – the presence of a potential continuing phonetic sequence allows for activation of the longer lexical item, despite the fact that there is a clear change in talker. This suggests that changes in talker are not, by themselves, sufficient to disrupt ongoing processes of lexical access.

Experiment 4

One final concern is that the cross-talker effects in Experiment 2 might be the result of some form of strategic effect. Although we had attempted to avoid most strategic effects through precise counterbalancing (for example, cross-voice items were followed equally often by real words as by nonwords, and cross-voice items were no more likely than single-voice items to be combinable into a larger disyllabic word), there was still a possibility that participants may have recognised either that a proportion of word pairs could be combined to make longer, disyllabic sequences, or that a substantial proportion of trials had a semantic relationship between the auditory word pair and the visual item. More specifically, one-fourth of the items in Experiment 2 could have been combined in that manner, and all of these items were followed by a real word; perhaps this was too large of a ratio, such that participants began learning this pattern and responding strategically on that basis. Although this proportion was modelled after that of other lexical decision studies (such as Gow & Gordon, 1995) the changes in talker might have highlighted the potential parsing ambiguity, such that participants grew aware of the possibility of disyllabic words over the course of the experiment. If participants were in a listening mode where they might expect to put words together across talkers, this could have led to a talker-mixing effect that does not accurately represent real-world listening conditions.

To avoid this concern, the current study attempts to replicate Experiment 2, but with a substantially reduced proportion of trials in which the two-word sequences can be combined into a larger, disyllabic word, and, likewise, a substantially reduced proportion of potentially related trials. In this study, only 20 trials (out of 400) had this situation in which a word spoken by a male talker and one spoken by a female talker could potentially be combined; this reduced likelihood is unlikely to have pushed listeners into an atypical mode of listening. If we nonetheless find cross-talker lexical access effects in this study, it would be much stronger evidence that the process of lexical access routinely activates sequences that crosstalker boundaries.

Method

Participants

Forty-eight members (34 female, 14 male) of the University of Maryland community took part in this experiment. All participants were right-handed, native speakers of English with no reported history of a speech, hearing, or reading disorder. Participants received course credit or a cash payment for their participation. The data from three additional participants were excluded for poor accuracy; data from one other were excluded because the participant was left-handed. This left a total of 12 participants hearing each of 4 lists. This is a larger number of participants than in Experiment 2, but this was deemed necessary to counteract the smaller number of target trials in this experiment (40 primed and 40 unprimed trials per participant, as compared to 120 of each in Experiment 2).

Stimuli

To ensure the generality of our cross-talker effects, all items in this experiment consisted of novel recordings, made by two new talkers (one male, one female). Other aspects of the recording process were identical to Experiment 2.

A subset of the target word pairs from Experiment 2 (40 pairs; see Appendix 3) were selected for use in this experiment. Two student research assistants, unaware of the purpose of the study, identified an additional 720 words to be recorded, which were then combined into 360 pairs. These research assistants also identified an additional set of words and nonwords to be presented visually as fillers.

This resulted in a total of 400 pairs of auditory words to be presented to listeners; half were followed by visual real words, and half by nonwords. As in Experiment 2, the auditory word pairs were both spoken by the female voice on 25% of the trials, and both by the male voice on 25%. Of the other trials, half had a female word followed by a male word, and half a male word followed by a female word.

However, unlike in Experiment 2, only 10% of the trials (40 out of 400) were ones in which the two-word sequence could be potentially combined to create a longer disyllabic word. Of these 40 target trials, 10 each occurred in a male–male voice combination, female–female voice combination, male–female voice combination and female–male voice combination. Thus, across the entire experiment, only 20 trials (out of 400) were both cross-voice and combinable; this low proportion makes it highly unlikely that participants would learn this relationship over the course of the study. The presence of 360 trials (90%) with uncombinable sequences (such as "oat-lease" and "back-luck") should avoid any implicit encouragement to treat auditory primes as single combined sequences.

Of the 400 auditory word pairs, 200 were followed by a visual nonword on the screen, and 200 by a real word. Of those 200 visual word trials, 40 were potentially primed by a 2-word related sequence; the other 160 real word items consisted of 40 trials in which those same target visual words were presented after an uncombinable auditory word pair (unrelated trials) and 120 trials that were considered fillers (where a visual nontarget was preceeded by two un-combinable auditory words). The 200 visual nonwords were always preceeded by 2 un-combinable auditory words.

Procedure

The procedure used was identical to that in Experiment 2, except that participants heard a total of 400 test trials in random order, with breaks at evenly spaced intervals throughout the experiment. In addition, this experiment was run using Psyscope X, rather than Psyscope (http://psy.ck.sissa.it/index.html). Participant RTs were measured from the onset of the visual item to the button press; RTs longer than 2000 ms or less than 50 ms were ignored (76 out of 19,200 trials, or 0.4%).

Results and discussion

Average lexical decision accuracy was 94.54%, demonstrating that participants were accurate at the task. As in Experiments 2 and 3, we conducted $2 \times 2 \times 2$ ANOVAs, both by subjects and by items; analyses were performed after removing outliers from the RT data (82 out of 3840 target trials, or 2.1%), and using arcsin transforms for accuracy. RT data are shown in Figure 4, and statistical results in Table 4.

The RT data showed a main effect of prime relatedness ($p_1 = .005$; $p_2 = .002$), with significantly faster responding in the related condition (533 ms vs. 547

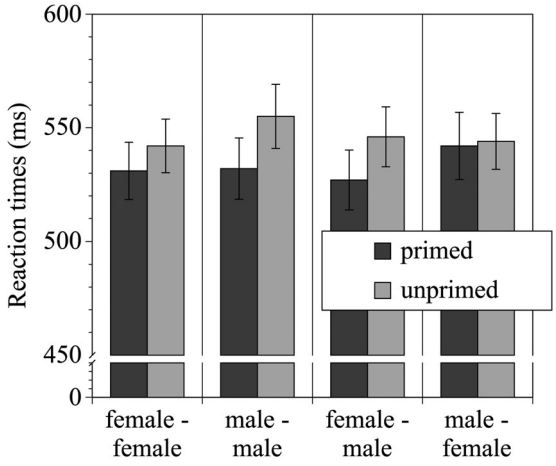

Figure 4. RTs from Experiment 4 to visual words preceded by word pairs that could either be combined into a related word (i.e. *two-lips* related to "flower"; primed) or which could not (unprimed), spoken either by a single voice or a voice combination. Error bars represent standard error.

ms). The size of the effect was smaller here than in the earlier experiments, although this could be in part the result of the much smaller number of trials (and resultant greater variability). Nonetheless, the effect remained significant by both subjects and items.

There was no main effect of the number of voices and only a marginal effect of the first voice. Critically, there were no significant interactions, although the three-way interaction was marginal in the subjects analysis only. The lack of any hint of an interaction between relatedness and the number of voices ($p_1 = .34$; $p_2 = .90$) suggests that listeners were no less likely to show a prime relatedness effect in the cross-voice conditions than in the single-voice conditions. However, looking at the RT data in the conditions individually, the relatedness effect was significant in only two cases: following a female–male cross-voice item (19 seconds, $t_1(47) = 2.63$, $p = .011$; $t_2(39) = 2.40$, $p = .02$), and following a male–male single-voice item by subjects (23 seconds, $t_1(47) = 2.54$, $p = .014$; $t_2(39) = 1.88$, $p = .067$). The trend was not significant following either the male–female cross-voice items ($t_1(47) = 0.13$, $p = .90$; $t_2(39) = 0.67$, $p = .51$), or the female–female single-voice items ($t_1(47) = 1.57$, $p = .12$; $t_2(39) = 1.23$, $p = .23$). Thus, despite the lack of an interaction, there is some hint that perhaps the effect is not equivalent in all cases. Still, it does not appear that the cross-voice items were behaving particularly differently than the single-voice items. If anything, the lack of an effect was tied to cases where the second voice was female, regardless of the first voice. While this does perhaps complicate the interpretation somewhat, there is still no evidence to suggest that a change in voice limits the activation of the longer potential word.

Accuracy data also showed a main effect of relatedness, although this was present only by subjects ($p_1 = .005$; $p_2 = .29$), with greater accuracy in the related condition. There were no other main effects and no interactions. Thus, based on both accuracy and RT measures, we find a general pattern of prime relatedness that is comparable in the single-voice and cross-voice conditions. This effect is stronger in the RT data than in the accuracy data, but is in the same direction in both, suggesting that it does not represent a strategic speed-accuracy trade-off. The effect is certainly much weaker in the current study than in the prior ones. But that such a priming effect occurred despite the very small number of related cross-voice trials suggests that this effect is not the result of atypical listening strategies employed by the participants, but may instead reflect a natural tendency to activate potential words in the lexicon despite the presence of talker changes and acoustic word boundary cues.

Table 4. Statistical results from Experiment 4.

Effect	By subjects	By items
RT data		
Prime relatedness	$F_1(1,47) = 8.64$, $p = .005$, $\eta_p^2 = .155$	$F_2(1,39) = 11.27$, $p = .002$, $\eta_p^2 = .224$
Number of voices	$F_1(1,47) = 0.01$, $p = .93$, $\eta_p^2 < .001$	$F_2(1,39) = 0.60$, $p = .44$, $\eta_p^2 = .015$
First voice	$F_1(1,47) = 2.97$, $p = .091$, $\eta_p^2 = .059$	$F_2(1,39) = 2.91$, $p = .096$, $\eta_p^2 = .069$
Relatedness × number of voices	$F_1(1,47) = 0.95$, $p = .34$, $\eta_p^2 = .02$	$F_2(1,39) = 0.02$, $p = .90$, $\eta_p^2 < .001$
Relatedness × first voice	$F_1(1,47) = 0.16$, $p = .69$, $\eta_p^2 = .003$	$F_2(1,39) = 0.18$, $p = .68$, $\eta_p^2 = .004$
Number of voices × first voice	$F_1(1,47) = 0.001$, $p = .98$, $\eta_p^2 < .001$	$F_2(1,39) = 0.001$, $p = .98$, $\eta_p^2 < .001$
Three-way interaction	$F_1(1,47) = 3.39$, $p = .072$, $\eta_p^2 = .067$	$F_2(1,39) = 0.99$, $p = .33$, $\eta_p^2 = .025$
Accuracy data		
Prime relatedness	$F_1(1,47) = 8.78$, $p = .005$, $\eta_p^2 = .157$	$F_2(1,39) = 1.15$, $p = .29$, $\eta_p^2 = .029$
Number of voices	$F_1(1,47) = 0.40$, $p = .84$, $\eta_p^2 = .001$	$F_2(1,39) = 0.12$, $p = .73$, $\eta_p^2 = .003$
First voice	$F_1(1,47) = 1.17$, $p = .29$, $\eta_p^2 = .024$	$F_2(1,39) = 3.86$, $p = .056$, $\eta_p^2 = .09$
Relatedness × number of voices	$F_1(1,47) = 1.07$, $p = .31$, $\eta_p^2 = .022$	$F_2(1,39) = 0.16$, $p = .69$, $\eta_p^2 = .004$
Relatedness × first voice	$F_1(1,47) = 0.39$, $p = .54$, $\eta_p^2 = .008$	$F_2(1,39) = 0.08$, $p = .78$, $\eta_p^2 = .002$
Number of voices × first voice	$F_1(1,47) = 1.42$, $p = .24$, $\eta_p^2 = .029$	$F_2(1,39) = 0.43$, $p = .52$, $\eta_p^2 = .011$
Three-way interaction	$F_1(1,47) = 0.19$, $p = .67$, $\eta_p^2 = .004$	$F_2(1,39) = 0.19$, $p = .67$, $\eta_p^2 = .005$

General discussion

Prior work has suggested that while spoken word recognition can be influenced by a wide array of acoustic cues signalling potential boundaries between words, these word boundary cues are not always sufficient to prevent lexical access from occurring. The present work

extends this prior literature by demonstrating that even the substantial acoustic discontinuity induced by a change in talker does not prevent activation of items that span that boundary.

Two different types of acoustic boundaries were assessed in the present experiments. These consisted of boundaries caused by a change in talker, and boundaries caused by the end of a word. The data regarding word boundaries replicate the pattern of results found by Gow and Gordon (1995): when listeners hear a two-word sequence such as *tree owes*, they activate not only those words, but also longer words containing those two syllables ("trios").

The primary finding of the present paper, however, has to do with the effect of talker changes. Surprisingly, listeners' lexical activation seemed relatively unaffected by a change in talker; the acoustic cues indicating a change in talker did not prevent lexical access from occurring. (Or, to put it another way, the presence of a cue to a change in talker identity did not terminate the ongoing process of lexical activation.) Listeners showed activation for the concept "neon" when hearing *knee* followed by *on* spoken by a single talker, and continued to do so when the two words were spoken by two different talkers. Yet they did not show such activation after hearing *knee* followed by a different word. This suggests that listeners consistently integrated portions of the speech signal spoken by different talkers, and accessed whatever lexical entries resulted from these amalgamations. This was the case despite the fact that the different talkers were extremely disparate acoustically; the gender difference between voices is quite obvious, such that participants could not have mistaken the two voices as having come from the same talker. Rather, it appears that cues to talker changes, like those to word boundaries, do not disrupt lexical access.

Moreover, the individual words in Experiments 2 and 4 were not only from two different talkers, but were produced in isolation and concatenated, such that there was no phonetic coherence or coarticulatory cues between the two syllables. It is hard to reconcile the lack of an effect of these abrupt acoustic changes with the notion of an explicit sub-lexical segmentation process. These findings suggest that rather than strategically considering multiple parsings only in those situations in which there is a true acoustic ambiguity, listeners will consider multiple parsings *despite* the presence of disambiguating acoustic information. This would seem to support models of lexical access in which segmentation into individual words is a *result* of lexical competition and identification, rather than being a precursor to it.

Although the distinction between lexical and acoustic-driven segmentation has often been treated as being a binary distinction, several researchers have suggested compromise positions (Mattys et al., 2005), in which acoustic cues serve to facilitate the lexical activation of some items over others. Yet, surprisingly, we not only found evidence of activation of the longer conjoined word, but also failed to find any apparent reduction in size of this activation when there was vs. was not a talker change. This might suggest that effects of acoustic cues (Lehiste, 1960; Nakatani & Dukes, 1977) take time to build up, and thus only have an influence later in processing, rather than prior to lexical access.

These findings also suggest that lexical representations are at least partially abstract with regards to talker information. Early episodic theories posited more exemplar-based lexicons (e.g. Goldinger, 1998), in which individuals have stored representations of words as spoken by a variety of different people. More recent theories suggest that both talker-specific/episodic and abstract information are stored, but may be accessed at different points in lexical processing or in different contextual situations (Mattys, Davis, Bradlow, & Scot, 2012; Mattys & Liss, 2008; McLennan, Luce, & Charles-Luce, 2003, 2005). More specifically, Luce, McLennan, and Charles-Luce (2003) have argued that "there is a time-course to … spoken word recognition, such that immediate processing is dominated by abstract codes, whereas specific information takes time to percolate through the system and have its effects on perception" (p. 198). Supporting this argument, McLennan and Luce (2005) found that the time frame in which people responded altered the degree to which they showed talker-specific effects: when a listening task was relatively easy, participants responded more quickly, and information regarding talker identity did not influence lexical access (as shown through repetition priming). However, when the authors forced participants to respond more slowly (by making the task more difficult), information about talker identity did influence lexical processing.

Results from the current study support this notion that at least the early stages of lexical access are based on relatively abstract representations. It is very unlikely that our listeners have previously experienced many words that were spoken by talkers of two different genders; listeners may have heard *toucan* produced by a variety of talkers, both male and female, but likely had not previously heard amalgamations of two voices. If lexical access were occurring primarily on the basis of talker-specific episodic traces, one might predict that the single-voice items would be more similar to participants' stored representations than would the cross-voice items, resulting in less activation for sequences

that crossed a talker change than for those that were within a single talker. Since we found no evidence for reduced activation in the cross-voice condition, it suggests that these items activated stored representations as strongly as did the single-voice sequences, in line with a more abstract representation. But since the stimuli in the current study were presented in excellent listening conditions (in a quiet room over headphones), with a relatively short time period between the offset of the auditory stimulus and the onset of the visual target, we may have encouraged listeners to respond too quickly for effects of talker identity to arise. Perhaps our effects would have been different if we had either introduced additional lag time between the auditory and visual stimuli (providing more time for indexical information to influence word recognition), or encouraged them to spend more time processing the auditory stimulus (such as by presenting it in noise). Unfortunately, such additional delay would also provide more opportunity for conscious strategic processing, and thus lead to a potential confound, which is why we have not tested this prediction explicitly.

Given that even the (quite drastic) changes in acoustic properties that accompany a change in talker gender proved insufficient to disrupt lexical access, what would be? We expect a pause would be, although the duration would need to be large enough to be clearly distinct from a voice onset time closure. Beyond this, however, it is unclear. One possibility is that while talker differences alone do not prevent lexical access, a change in language (as in a code-switched utterance), or a combination of talker differences and language differences might; thus, a male voice saying a word in one language followed by a female voice saying a word in a second might not lead to a combined percept. This could be a topic for future research. Another likely possibility is that when there are multiple talkers speaking *simultaneously*, lexical access would occur only within one stream of sound. For example, imagine a situation in which a male voice said the two words, *kid bat*, while a female voice said the two words, *two knee*. In this situation, there is again the potential for activation of a longer, cross-talker sequence ("kidney", formed from the first word in the male voice, and the second word in the female voice). However, unlike in the present study, there is a competing perceptual grouping that might encourage segregation of the two voices into distinct streams. Such situations, in which two talkers' productions overlap in time, are likely to be quite common outside of the laboratory; despite this, we predict that illusory conjunctions would be infrequent because they would involve not only combining across talkers, but also combining across separate streams of speech. Yet prior work has suggested that illusory conjunctions can indeed occur; Mattys and Samuel (1997) found that listeners heard illusory conjunctions in a dichotic listening task (e.g. hearing "controversy" when presented with *kintroversy* in one ear and *bosglorafe* in the other). This suggests that lexical activation may not be limited to a single stream, even when there is clear evidence for auditory scene analysis.

To some degree, the types of sequential cross-talker conjunctions examined here might be thought to be unlikely to occur outside the laboratory. Even when one talker finishes another's sentence, the time delay between one speakers' conclusion and another's onset is unlikely to be as short as that found here. Some recent evidence contradicts this intuition, however. de Ruiter, Mitterer, and Enfield (2006) analysed the average time between the end of one speaker's turn and the start of the next speaker's turn in a Dutch telephone conversation database, which they referred to as *floor transfer offsets*, or FTOs. They found that 45% of all speaker transitions had an FTO of "between −250 and 250 ms". Similarly, Stivers et al. (2009) compared data from videorecordings of conversations in 10 different languages; in general, the modal temporal offset between the end of a speech turn and its response was between 0 and +200 ms, with 4 languages having mean values of 110 ms or less. These results suggest that the 100 ms offset used here may not, in fact, be that uncommon, and listeners may frequently hear speech in which different talkers' words occur in close proximity. This may also be more common for some listeners than for others; perhaps listeners who experience this situation more often would be more likely to limit activation to information from a single talker. Understanding the factors that influence lexical access in these multi-talker environments is important for our more general understanding of spoken language processing particularly in complex multi-talker listening environments.

In conclusion, the present study examined whether information from multiple talkers would be combined during the process of lexical access. Results suggest that acoustic cues indicating the presence of multiple talkers are not sufficient to disrupt lexical access, even when combined with clear word boundary cues. This suggests that multiple lexical hypotheses may be entertained whenever consistent input in the signal is encountered, and that acoustic cues to talker identity are not sufficient to block activation of these multiple interpretations. This supports models in which lexical segmentation occurs as part of the multi-dimensional constraint-satisfaction process of identifying words, rather than being a distinct stage of prelexical processing.

Acknowledgements

The author thanks Cheryl Tarbous, Matt Winn, Amy Robbins, Ryan Hurm, Jenni Zobler, Catherine Wu, Chris Heffner, Krista Voelmle, Andrea Fisher and David Rossell for assistance in stimulus creation; David Gow, Jim Mullennix, Chris Heffner and James Sawusch for several helpful discussions; and the following students for assistance in scheduling and testing participants: Alison Arnold, Sarah Aylor, Amelie Bail, Catherine Bender, Taryn Bipat, Devon Brunson, Rachel Childress, Emilie Clingerman, Alyssa Cook, Jennifer Coon, Nicole Craver, Hayley Derris, Justine Dombroski, Sara Dougherty, Lauren Evans, Annie Ferruggiaro, Lyana Kardanova Frantz, Nikki Friedman, Katherine Gagan, Arielle Gandee, Sarah Haszko, Devin Heit, Megan Janssen Crenshaw, Lacey Kahner, Caroline Kettl, Hannah Kim, Penina Kozlovsky, Stephanie Lee, Rachel Lieberman, Amanda Lin, Perri Lieberman, Heather McIntosh, Giovanna Morini, Vidda Moussavi, Molly Nasuta, Jessica Nwaogbe, Maura O'Fallon, Amanda Pasquarella, Jessica Pecora, Mariah Pranger, Rachel Rhodes, Amy Robbins, Allie Rodriguez, Toni Rodriguez, Katerina Sanders, Kate Shapiro, Rebecca Sherman, Lauren Simpson, Emily Singer, Emily Slonecker, Veronica Son, Rebecca Spencer, Cheryl Tarbous, Ashley Thomas, Nicole Tobin, Julian Vesnovsky, Krista Voelmle, Donnia Zack-Williams, Kimmie Wilson and Jenni Zobler.

References

van Alphen, P. M., & van Berkum, J. J. A. (2010). Is there pain in champagne? Semantic involvement of words within words during sense-making. *Journal of Cognitive Neuroscience*, *22* (11), 2618–2626. doi:10.1162/jocn.2009.21336

Bowers, J. S., Davis, C. J., Mattys, S. L., Damian, M. F., & Hanley, D. (2009). The activation of embedded words in spoken word identification is robust but constrained: Evidence from the picture-word interference paradigm. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(5), 1585–1597. doi:10.1037/a0015870

Christophe, A., Gout, A., Peperkamp, S., & Morgan, J. (2003). Discovering words in the continuous speech stream: The role of prosody. *Journal of Phonetics*, *31*(3/4), 585–598. doi:10.1016/S0095-4470(03)00040-8

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers*, *25*(2), 257–271. doi:10.3758/BF03204507

Cole, R. A., & Jakimik, J. (1980). A model of speech perception. In R. A. Cole (Ed.), *Perception and production of fluent speech* (pp. 133–163). Hillsdale, NJ: Erlbaum.

Davis, M. H., Marslen-Wilson, W. D., & Gaskell, M. G. (2002). Leading up the lexical garden path: Segmentation and ambiguity in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(1), 218–244. doi:10.1037/0096-1523.28.1.218

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*(2), 251–279. doi:10.1037/0033-295X.105.2.251

Gow, D. W.Jr., & Gordon, P. C. (1995). Lexical and prelexical influences on word segmentation: Evidence from priming. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(2), 344–359. doi:10.1037/0096-1523.21.2.344

Kim, D., Stephens, J. D. W., & Pitt, M. A. (2012). How does context play *a part* in splitting words *apart*? Production and perception of word boundaries in casual speech. *Journal of Memory & Language*, *66*(4), 509–529. doi:10.1016/j.jml.2011.12.007

Klatt, D. H. (1980). Speech perception: A model of acoustic–phonetic analysis and lexical access. In R. A. Cole (Ed.), *Perception and production of fluent speech* (pp. 243–288). Hillsdale, NJ: Lawrence Erlbaum.

Lehiste, I. (1960). *An acoustic–phonetic study of internal open juncture*. New York. doi:10.1159/000258062

Luce, P. A. (1986). A computational analysis of uniqueness points in auditory word recognition. *Perception & Psychophysics*, *39*, 409–420. doi:10.3758/BF03212485

Luce, P. A., & Cluff, M. S. (1998). Delayed commitment in spoken word recognition: Evidence from cross-modal priming. *Perception & Psychophysics*, *60*(3), 484–490. doi:10.3758/BF03206868

Luce, P. A., Goldinger, S. D., Auer, E. T., & Vitevitch, M. S. (2000). Phonetic priming, neighborhood activation, and PARSYN. *Perception & Psychophysics*, *62*, 615–625.

Luce, P. A., & Lyons, E. A. (1999). Processing lexically embedded spoken words. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(1), 174–183. doi:10.1037/0096-1523.25.1.174

Luce, P. A., McLennan, C. T., & Charles-Luce, J. (2003). Abstractness and specificity in spoken word recognition: Indexical and allophonic variability in long-term repetition priming. In J. S. Bowers & C. J. Marsolek (Eds.), *Rethinking implicit memory* (pp. 197–214). Oxford: Oxford University Press.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*, 71–102. doi:10.1016/0010-0277(87)90005-9

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, *10*, 29–63. doi:10.1016/0010-0285(78)90018-X

Mattys, S. L. (2004). Stress versus coarticulation: Towards an integrated approach to explicit speech segmentation. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(2), 397–408. doi:10.1037/0096-1523.30.2.397

Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scot, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes*, *27*, 953–978. doi:10.1080/01690965.2012.705006

Mattys, S. L., & Liss, J. M. (2008). On building models of spoken-word recognition: When there is as much to learn from natural "oddities" as artificial normality. *Perception & Psychophysics*, *70*(7), 1235–1242. doi:10.3758/PP.70.7.1235

Mattys, S. L., & Samuel, A. G. (1997). How lexical stress affects speech segmentation and interactivity: Evidence from the migration paradigm. *Journal of Memory & Language*, *36*, 87–116. doi:10.1006/jmla.1996.2472

Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General*, *134*(4), 477–500. doi:10.1037/0096-3445.134.4.477

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

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*(2), 306–321. doi:10.1037/0278-7393.31.2.306

McLennan, C. T., Luce, P. A., & Charles-Luce, J. (2003). Representation of lexical form. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *29*(4), 539–553. doi:10.1037/0278-7393.29.4.539

McLennan, C. T., Luce, P. A., & Charles-Luce, J. (2005). Representation of lexical form: Evidence from studies of sub-lexical ambiguity. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(6), 1308–1314. doi:10.1037/0096-1523.31.6.1308

McQueen, J. M. (2005). Speech perception. In K. Lamberts & R. Goldstone (Eds.), *The handbook of cognition* (pp. 255–275). London: Sage. doi:10.4135/9781848608177.n11

McQueen, J. M., Cutler, A., Briscoe, T., & Norris, D. (1985). Models of continuous speech recognition and the contents of the vocabulary. *Language and Cognitive Processes*, *10*, 309–331. doi:10.1080/01690969508407098

Nakatani, L. H., & Dukes, K. D. (1977). Locus of segmental cues for word juncture. *Journal of the Acoustical Society of America*, *62*(3), 714–719. doi:10.1121/1.381583

Newman, R. S., Sawusch, J. R., & Wunnenberg, T. (2011). Cues and cue interactions in segmenting words in fluent speech. *Journal of Memory & Language*, *64*, 460–476. doi:10.1016/j.jml.2010.11.004

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

Norris, D., Cutler, A., McQueen, J. M., & Butterfield, S. (2006). Phonological and conceptual activation in speech comprehension. *Cognitive Psychology*, *53*, 146–193. doi:10.1016/j.cogpsych.2006.03.001

Norris, D., McQueen, J. M., Cutler, A., & Butterfield, S. (1997). The possible-word constraint in the segmentation of continuous speech. *Cognitive Psychology*, *34*, 191–243. doi:10.1006/cogp.1997.0671

Quené, H. (1992). Durational cues for word segmentation in Dutch. *Journal of Phonetics*, *20*, 331–350. doi:10.1121/1.407504

Reddy, R. (1976). Speech recognition by machine: A review. *Proceedings of the IEEE*, *64*, 501–531. doi:10.1109/PROC.1976.10158

de Ruiter, J. P., Mitterer, H., & Enfield, N. J. (2006). Projecting the end of a speaker's turn: A cognitive cornerstone of conversation. *Language*, *82*(3), 515–535. doi:10.1353/lan.2006.0130

Shillcock, R. (1990). Lexical hypotheses in continuous speech. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 24–49). Cambridge, MA: Bradford.

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., … Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(26), 10587–10592. doi:10.1073/pnas.0903616106

Tabossi, P., Burani, C., & Scott, D. (1995). Word identification in fluent speech. *Journal of Memory & Language*, *34*, 440–467. doi:10.1006/jmla.1995.1020

Vitevitch, M. S. (2003). Change deafness: The inability to detect changes between two voices. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(2), 333–342. doi:10.1037/0096-1523.29.2.333

Vroomen, J., & de Gelder, B. (1997). Activation of embedded words in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *23*(3), 710–720. doi:10.1037/0096-1523.23.3.710

Wester, M. (2012). Talker discrimination across languages. *Speech Communication*, *54*, 781–790. doi:10.1016/j.specom.2012.01.006

Zwitserlood, P. (1989). The locus of the effects of sentential-semantic context in spoken-word processing. *Cognition*, *32*, 25–64. doi:10.1016/0010-0277(89)90013-9

Appendix 1. Items used in Experiment 1. The first item is the auditory prime with a slash indicating the location of the cross-splice; the second item is the visual lexical decision target.

A/dult – child
A/pple – orange
Ar/my – navy
Ar/row – bow
Au/tumn – leaves
Bu/cket – pail
Bu/tter – bread
Cac/tus – desert
Can/dle – wax
Car/pet – rug
Cei/ling – floor
Christ/mas – tree
Cir/cle – square
Co/rrect – wrong
Daugh/ter – son
Dia/mond – ring
Dir/ty – clean
Doc/tor – nurse
Eas/ter – bunny
E/vil – bad
Fea/ther – light
Fo/rest – trees
Ha/mmer – nail
Ha/ppy – sad
Hel/lo – goodbye
Hus/band – wife
I/nner – outer
In/sect – bug
Ke/tchup – mustard
Ki/tty – doggy
La/dies – gentlemen
Mar/riage – wedding
Mill/ion – dollar
Mo/ther – father

Nee/dle – thread
Num/ber – letter
O/pen – close
Out/let – plug
O/ver – under
Pa/per – pencil
Pe/pper – salt
Pol/ice – officer
Po/ny – horse
Pop/corn – movie
Pri/son – jail
Pump/kin – halloween
Ques/tion – answer
Ro/bber – thief
Ru/by – red
Sal/sa – chips
Sil/ver – gold
Sis/ter – brother
Slee/py – tired
Suc/ceed – fail
Su/mmer – winter
Ta/ble – chair
Thun/der – lightening
Ti/ny – small
To/day – tomorrow
Un/cle – aunt

Appendix 2. Items used in Experiment 2. The first two words are the two-word version of the prime; the final item is the visual lexical decision target.

A choir – get
A count – bank
A cross – down
A cyst – help
A muse – laugh
A salt – battery
A tack – fight
Add dress – street
Air row – bow
Apart meant – building
Are me – navy
Axe sent – foreign
Bay be – infant
Con vent – nun
Be tray – friend
Can dull – flame
Can new – oar
Can sir – disease
Can teen – water
Can't elope – fruit
Cap size – sink
Cap tin – ship
Car go – ship
Car pet – rug
Car tells – drugs
Car tunes – television
Cash you – nut
Chilled wren – kids
Chris miss – holiday
Claws it – clothes
Con test – winner
Core wrecked – wrong
Cough inn – death
Cry sis – problem
Cull loan – smell
Depart meant – store
Dew owe – two
Die sect – frog
Doll fins – Miami
Drag inn – fire
Eggs it – sign
Fan sea – dress
For words – back
Four get – remember
Gore may – food
Guard in – flowers
Hair ring – fish
Hay low – angel
Hell met – football
Here row – villain
High gene – clean
Hiss story – past
Honey do – melon
Hue man – being
Inn jury – hurt
Inn turn – student
Inn vest – money
Jack kit – coat
Kid knee – beans
Knee on – lights
Less inn – learn
Let us – salad
Mare ridge – wedding
May hem – chaos
May tricks – movie
Men you – food
Mill do – moldy
Miss stake – wrong
My great – geese
Nap kin – messy
Oar kid – flower
Off in – sometimes

Out let – plug
Owe bay – rules
Pair rents – children
Pan tree – food
Pass port – travel
Pass tell – colors
Pay per – pencil
Per fume – scent
Per pull – pink
Per son – man
Pig meant – color
Pill low – sleep
Poll tree – chicken
Prod duct – buy
Prom miss – keep
Raise inn – grape
Ran dumb – weird
Ray on – fabric
Sadder day – Sunday
Sand witch – bread
Say lean – solution
Seas inn – winter
Sell fish – me
Sew low – alone
Sigh lent – night
Sin tax – words
Sing gull – married
Sir prize – party
Sir round – sound
Spare row – bird
State meant – bank
Term might – bug
Toy let – bathroom
Tree owe – three
Two can – bird
Two lips – pretty
Two pay – hair
Universe city – college
Wall let – money
Well come – mat
Were ship – praise
Will low – tree
Win dough – pane
Yell low – sun
You're inn – yellow

Appendix 3. Items used in Experiment 4. The first two words are the two-word version of the prime; the final item is the visual lexical decision target.

Ray on – fabric
Cap tin – ship
Mare ridge – wedding
Cull loan – perfume
Bee tray – friend
Sin tax – words
High gene – clean
Can sir – disease
Pair rents – children
Honey do – melon
Per pull – pink
Seas inn – winter
Pill low – sleep
Sing gull – married
A cyst – help
Are me – navy
Depart meant – store
Hue man – being
Well come – mat
Sand witch – bread
Miss steak – wrong
Car tunes – television
Cap size – sink
Kid knee – beans
Win dough – pane
Jack kit – coat
Con vent – nun
Out let – plug
Bay be – infant
Hun dread – dollars
Too pay – hair
Gore may – food
Poll tree – chicken
Ran dumb – weird
My great – geese
Hiss story – past
Universe city – college
Cry sis – problem
Prom miss – keep