Cognition 151 (2016) 96–107

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

What you see isn't always what you get: Auditory word signals trump consciously perceived words in lexical access

Rachel Ostrand [a,*], Sheila E. Blumstein [b], Victor S. Ferreira [c], James L. Morgan [b]

[a] Department of Cognitive Science, University of California, San Diego, United States
[b] Department of Cognitive, Linguistic, and Psychological Sciences, Brown University, United States
[c] Department of Psychology, University of California, San Diego, United States

ARTICLE INFO

Article history:
Received 3 January 2015
Revised 17 February 2016
Accepted 27 February 2016
Available online 21 March 2016

Keywords:
Auditory–visual integration
Lexical access
McGurk effect
Multisensory perception

ABSTRACT

Human speech perception often includes both an auditory and visual component. A conflict in these signals can result in the McGurk illusion, in which the listener perceives a fusion of the two streams, implying that information from both has been integrated. We report two experiments investigating whether auditory–visual integration of speech occurs before or after lexical access, and whether the visual signal influences lexical access at all. Subjects were presented with McGurk or Congruent primes and performed a lexical decision task on related or unrelated targets. Although subjects perceived the McGurk illusion, McGurk and Congruent primes with matching real-word auditory signals equivalently primed targets that were semantically related to the auditory signal, but not targets related to the McGurk percept. We conclude that the time course of auditory–visual integration is dependent on the lexicality of the auditory and visual input signals, and that listeners can lexically access one word and yet consciously perceive another.

1. Introduction

Speech comprehension is a complex, multi-staged process. Although speech perception is primarily driven by the auditory signal (Barutchu, Crewther, Kiely, Murphy, & Crewther, 2008; Erber, 1975), visual information, such as that provided by mouth movements, can have an influence as well (Fort, Spinelli, Savariaux, & Kandel, 2010; Green, 1998; Summerfield, 1987), especially in noisy or degraded environments (Erber, 1975; Grant & Seitz, 2000; Sumby & Pollack, 1954). This implies that auditory and visual signals are integrated into a single representation at some point during processing. The present work addresses whether auditory–visual (AV) integration occurs before or after the component stimuli access the lexical-semantic network, and thus what role (if any) visual speech information plays in lexical access.

McGurk and MacDonald (1976) first reported the McGurk Effect, in which mismatching auditory and visual signals perceptually combine. The result is that listeners consciously perceive a stimulus which is a fusion of the auditory and visual inputs, and thus is different from what would be perceived by hearing the auditory signal alone. To create these integrated auditory–visual percepts, a video of a speaker mouthing a stimulus is dubbed with an auditory track differing by one consonant's place of articulation. People often report perceiving McGurk stimuli as a fusion of phonetic features from the auditory and visual signals. For example, auditory [ba] paired with visual [ga] or [da] is generally consciously perceived as *da*. This effect is remarkable because of its illusory status – listeners report perceiving tokens that are distinct from the auditory signal, even though the auditory input is perceptually clear.[1]

A visual signal can especially affect the perception of a degraded auditory signal. In addition to having a stronger influence in noisy environments, there is some evidence from perceptual identification tasks that subjects perceive the McGurk illusion more frequently when the auditory signal is less "good" than the integrated signal. For example, two previous studies (Barutchu et al., 2008; Brancazio, 2004) provide evidence for a lexical bias in auditory–visual integration. Subjects were shown incongruent auditory–visual stimuli, and reported perceiving the fused AV

[1] In the literature, the "McGurk Effect" is sometimes used to refer more narrowly to instances in which the resulting percept corresponds neither to the auditory nor the visual stimulus but rather to something between the two. It is also sometimes used more broadly to refer to instances in which the visual stimulus causes perceivers to report something other than what they would report if exposed to the auditory signal alone. In this article, we use "McGurk Effect" in its broader sense.

* Corresponding author at: Department of Cognitive Science, 9500 Gilman Drive, #0515, La Jolla, CA 92093-0515, United States.
E-mail address: rostrand@cogsci.ucsd.edu (R. Ostrand).

http://dx.doi.org/10.1016/j.cognition.2016.02.019
0010-0277/

percept more often when the auditory signal was a nonword (e.g., *besk*) than when the auditory signal was a real word (e.g., *beg*). Similarly, subjects reported perceiving the fused percept more often when the visual signal (and fused percept) was a word (e.g., *desk*) than when it was a nonword (e.g., *deg*; but see Sams, Manninen, Surakka, Helin, & Kättö, 1998, for conflicting results). Thus, visual information seems to affect perception when it enhances access to the lexicon, by increasing the likelihood that a nonword auditory signal is comprehended as a real word.

These studies suggest that lexical characteristics of the auditory and visual signals affect whether the visual signal influences the outcome of conscious perception – what listeners report they heard. However, it remains unknown whether the same lexical characteristics of the auditory and visual signals influence lexical access – that is, the extent to which a given stimulus input activates lexical representations and information in the associated semantic network. The present work addresses this second question, specifically investigating the timing of AV-integration, and the situations in which each sensory signal does or does not influence lexical access.

Our research speaks to an ongoing debate as to whether AV-integration is an early or late process – pre- or post-lexical – and thus whether the combined percept or the separate sensory signals drive lexical access. There is some evidence for both possibilities; however, much of the previous work used nonword syllables (such as [ba]) rather than real words, making it impossible to draw conclusions about the time course of AV-integration relative to lexical access specifically. The current work challenges the assumption of a strict pre- or post-lexical dichotomy and considers alternative points at which the auditory and visual signals could be integrated.

The dominant view is that AV-integration is a strictly early, pre-lexical process; that is, that the separate auditory and visual inputs are fused into a single stimulus before lexical access occurs (see Massaro & Jesse, 2007, for a discussion). In this case, the integrated McGurk percept – not the auditory signal – is the lookup key in the lexicon, accessing its own lexical-semantic entry and associates. This would imply that AV-integration operates in a purely bottom-up direction, and occurs similarly regardless of the lexicality or other characteristics of either sensory input signal. Supporting early integration, Sams et al. (1998) found that subjects were equally likely to fuse conflicting auditory and visual streams into nonwords as into words. This was true regardless of whether stimuli were isolated words or were predictable from the preceding sentence context, suggesting that AV-integration occurred before (and irrespective of) word identification (though note, as discussed below, Brancazio, 2004 reports different results). Additionally, some neuropsychological evidence suggests that AV-integration is an early process. Colin et al. (2002) exposed subjects to a high proportion of congruent AV stimuli (e.g., $bi_{Aud}bi_{Vis}$), interspersed with a few incongruent AV stimuli (e.g., $bi_{Aud}di_{Vis}$). The incongruent stimuli elicited a mismatch negativity (MMN), an automatic and pre-attentive electroencephalography (EEG) component. However, infrequent visual-only stimuli (e.g., $Ø_{Aud}di_{Vis}$ presented interspersed with frequent $Ø_{Aud}bi_{Vis}$) elicited no MMN. As infrequent visual stimuli seem not to elicit an MMN, the differing visual signals of the incongruent and congruent AV stimuli could not have triggered the observed MMN component, and thus subjects must have integrated the auditory and visual streams of the McGurk stimuli. And because the incongruent AV items elicited an MMN even though the auditory signal was identical to that of the congruent items (both bi_{Aud}), AV-integration must have occurred early in processing (before the MMN occurred), and the MMN must have reflected the integrated AV percept (see also Besle, Fort, Delpuech, & Giard, 2004; Colin, Radeau, Soquet, & Deltenre, 2004; Saint-Amour, De Sanctis, Molholm, Ritter, & Foxe, 2007; Soto-Faraco, Navarra, & Alsius, 2004). This view is also supported by models of AV-integration such as the Fuzzy Logical Model of Perception (FLMP; Massaro, 1987) or Pre-labeling Model (Braida, 1991). For example, the FLMP predicts that although the differing sensory signals are initially evaluated separately, they are integrated prior to perception and interpretation of the stimulus.

In contrast, other research suggests that AV-integration is a late, post-lexical process, and thus lexical access should occur based on the information in (one or both) separate, un-integrated sensory input signals. As the auditory stream is usually more informative about speech than lip-reading is, under this account, the activated lexical item should derive from the auditory stimulus. Only later, after lexical access, would AV-integration occur, producing the perceptual experience of the McGurk Effect. Thus the combined percept, and the word or nonword it forms, would not (initially) contact the lexicon. Unlike the early integration account, under the post-lexical (late AV-integration) account, AV-integration does *not* occur irrespective of the input signals' properties. In this case, some incongruent stimuli may never get integrated, or may take longer to do so. Supporting the late time course, Brancazio (2004) and Barutchu et al. (2008) found a lexical bias in the McGurk effect. They found that AV stimuli with nonword auditory signals were more likely to be perceived as the McGurk fusion than those with word auditory signals, implying that lexical access had already occurred on the unimodal auditory signal to determine its lexicality. (Note, however, that these results are inconsistent with the results of Sams et al., 1998; Brancazio, 2004 suggests that this is due to shortcomings of the McGurk stimuli used by Sams et al.)

Similarly, a late integration account raises the possibility that top-down factors and semantic knowledge might influence whether listeners perceive the McGurk Effect at all. Indeed, listeners report more McGurk illusions and rate their perception closer to the fused word when the AV fusion is semantically congruent with a preceding sentence (Windmann, 2004), suggesting that listeners have access to the meaning of the sensory input signals (and their semantic associates) before integrating them (or not). Models of AV-integration such as the Post-labeling Model (Braida, 1991) support a late time course of integration.

The results from the experiments reported here suggest that the time course of AV-integration is more nuanced than a strict binary choice between pre-lexical or post-lexical integration. We propose a third possibility – that the time course, and the likelihood of success, of AV-integration is dependent on the lexicality of the two input signals. Whether AV-integration occurs before or after lexical access could depend on properties of the specific auditory and visual inputs, rather than having a fixed time course. There are some hints in prior research supporting this hypothesis. For example, Baart and Samuel (2015) presented subjects with spoken words and nonwords that differed at the onset of the third syllable (like "banana" and "banaba"). Additionally, the third syllable was either presented auditory-only, visual-only (i.e., mouthed), or auditory–visual. They found that both lexical status and presentation modality modulated subjects' ERP activity. However, the two factors did not affect each other's degree of influence, and occurred at the same time points. Although Baart and Samuel (2015) did not test incongruent AV stimuli, their results suggest that lexical access and the integration of auditory and visual signals might, in certain circumstances, occur in parallel.

In the present experiments, subjects performed lexical decisions on auditory target items that were semantically related or unrelated to preceding auditory–visual primes. The primes were either created from mismatching AV signals (McGurk) or matching AV signals (congruent controls). This priming task allows for the detection of words that the auditory–visual prime stimuli activate in the lexicon. In Experiment 1, for each McGurk prime, either the auditory signal or the integrated auditory–visual (McGurk) percept was a word; the other was a nonword. Congruent primes paired

each McGurk auditory signal with its matching visual signal, and provided a reaction time reference point for lexical access in the absence of integration of conflicting signals. If AV-integration is pre-lexical (and thus the fused AV stimulus is used for lexical access), then targets following primes which are perceived as real words should be responded to more quickly, and show a greater priming effect, than targets following primes which are perceived as nonwords. In contrast, if AV-integration is post-lexical (and thus the auditory signal is used for lexical access), then targets following primes which have real word auditory signals should be responded to more quickly, and show a greater priming effect, than targets following primes with nonword auditory signals.

In Experiment 2, all McGurk primes had both auditory and visual signals that were (different) real words; thus, both the auditory signal and the McGurk percept had an entry in the lexicon. Targets were semantically related to either the auditory signal or McGurk percept. If one group of target words is primed more, this would demonstrate that the corresponding part of the McGurk stimulus (auditory signal or integrated percept) accessed the lexicon.

2. Experiment 1

Experiment 1 was designed to adjudicate between the possibilities of pre-lexical and post-lexical auditory–visual integration. To do so, it employed McGurk primes that had either a real-word auditory signal and a nonword perceptual fusion, or a nonword auditory signal and a real-word perceptual fusion. Corresponding Congruent primes that matched each McGurk prime's auditory signal were also used. If primes with auditory-word signals (both McGurk and Congruent) elicit faster target responses and a larger priming effect (that is, related targets are responded to faster than unrelated targets) compared to their auditory-nonword counterparts, this would suggest that lexical access occurs based on the auditory stream and thus AV-integration occurs post-lexically. Note that in this case, targets following McGurk and Congruent primes should elicit the same reaction times as each other: As lexical access would depend solely on the auditory signal, the congruency of the visual signal should not affect lexical access, and thus reaction times. In contrast, if McGurk primes with auditory-nonword signals lead to faster target responses and a larger priming effect (compared to their auditory-word counterparts), this would suggest lexical access of the combined percept, and thus AV-integration occurs pre-lexically. In this case, targets following McGurk and Congruent primes should elicit *different* reaction times from each other. Target response times following McGurk auditory-nonwords should be faster than even their Congruent counterparts, because these McGurk primes are perceived as real words, whereas the corresponding Congruent primes are perceived as nonwords. Therefore, the pre- and post-lexical integration accounts make different predictions not only about the responses to auditory-word compared to auditory-nonword primes, but also as a function of each prime stimulus's congruency.

2.1. Method

2.1.1. Participants

Twenty-four Brown University students (13 male; 18–22 years old) participated. Two subjects' data were replaced due to instrument malfunction. All were native, monolingual English speakers.

2.1.2. Materials

Stimuli consisted of an auditory–visual prime followed, after a 50 ms inter-stimulus interval (ISI), by an auditory-only target. McGurk primes were composed of auditory and visual signals differing only on the initial consonant, and either the auditory signal or the McGurk percept was a real English word; the other was a nonword. In particular, the McGurk stimuli used in this experiment consisted of auditory and visual signals which were identical except for the place of articulation of the onset consonant (e.g., auditory "beef" combined with visual "deef"). This feature is often difficult to distinguish using auditory information alone (Miller & Nicely, 1955), and the visual signal can be used to disambiguate between possible consonants. When stimuli of this nature are combined, people generally report an illusory percept that matches the visual signal; thus, the intended McGurk percept matched the visual signal "deef". Congruent primes consisted of matching auditory and visual signals; half were real English words and half were nonwords.

In a pilot experiment, subjects watched AV stimuli of a woman speaking and rated, on a 5-point scale, the "goodness" of a queried onset consonant, either of the auditory signal or the intended integrated percept that matched the visual signal (where 1 = "poor example" and 5 = "excellent and clear example"). The 48 items with the highest McGurk Effect ratings (out of 117 McGurk stimuli presented in the pilot experiment) were chosen as stimuli for the main experiment; among these, word and nonword McGurk items were rated equally highly (McGurk-word: $M = 3.47$, McGurk-nonword: $M = 3.22$, $t(23) = 1.18$, $p = .249$). For both groups of McGurk items, ratings of the integrated McGurk percept ($M = 3.35$) were significantly higher than ratings of the auditory signal ($M = 2.42$; $t(47) = 4.83$, $p < .0001$), demonstrating that the selected stimuli successfully induced the McGurk Effect in subjects.

Each McGurk item had a Congruent counterpart which paired the McGurk item's auditory signal with its matching visual signal. For example, the McGurk item $beef_{Aud}deef_{Vis} \rightarrow deef_{Percept}$ had the corresponding Congruent item $beef_{Aud}beef_{Vis} \rightarrow beef_{Percept}$. Twenty-four McGurk primes had word auditory signals and nonword visual signals (e.g., $beef_{Aud}deef_{Vis} \rightarrow deef_{Percept}$); and thus there also were 24 Congruent word primes. The remaining 24 McGurk primes had nonword auditory signals and word visual signals (e.g., $bamp_{Aud}damp_{Vis} \rightarrow damp_{Percept}$), and 24 associated Congruent nonword primes (e.g., $bamp_{Aud}bamp_{Vis} \rightarrow bamp_{Percept}$). See Table 1 for sample prime and target stimuli and Appendix A for the complete stimuli list.

Each prime was paired with four auditory-only targets: two words (one semantically related, one unrelated), and two nonwords. Related targets were selected from the University of South Florida Free Association Norms database (Nelson, McEvoy, & Schreiber, 1998), the Edinburgh Associative Thesaurus (Kiss, Armstrong, Milroy, & Piper, 1973), or, when not available in either database, suggested by lab members. One- and two-syllable nonword targets were chosen from the ARC nonword database (Rastle, Harrington, & Coltheart, 2002).

2.2. Procedure

Participants watched trials on a monitor while listening through noise-attenuating headphones. On each trial, the AV prime appeared, played, and disappeared; there was a 50 ms ISI, followed by the auditory-only target. Participants were instructed to make a lexical decision as quickly as possible on the second item by pressing buttons labeled "word" or "nonword" on a button box. Button assignment was counterbalanced across subjects.

For each critical prime item, each participant was assigned either the two word targets or the two nonword targets. They heard each assigned target twice, once paired with the relevant McGurk prime (e.g., $bamp_{Aud}damp_{Vis}$–*wet*) and once with the corresponding Congruent prime (e.g., $bamp_{Aud}bamp_{Vis}$–*wet*; see Fig. 1). This allowed for the within-subject and within-item

Table 1
Sample Stimuli for Experiment 1. Congruent primes had matching auditory and visual signals; McGurk primes had mismatching auditory and visual signals. Targets were auditory-only.

Prime auditory lexicality	Prime congruency	Auditory signal	Visual signal	Perceived signal	Related target	Unrelated target
Nonword	Congruent	Bamp	Bamp	Bamp	Wet	Middle
	McGurk	Bamp	Damp	Damp		
Word	Congruent	Beef	Beef	Beef	Meat	Ask
	McGurk	Beef	Deef	Deef		

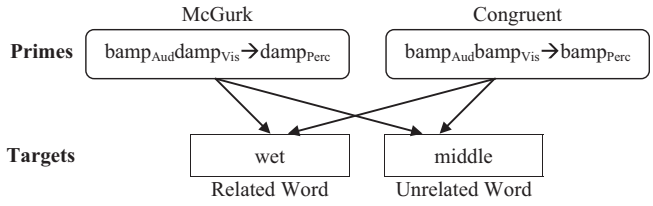

Fig. 1. One prime-target pairing in Experiment 1.

comparison of reaction times to a given target word following each prime condition. Stimuli were presented in two blocks separated by a participant-controlled break; the two repeated prime presentations and the two repeated target presentations were separated between blocks. Order of presentation of the two repeated primes and two repeated targets was counterbalanced across subjects. For example, one subject received bamp$_{Aud}$bamp$_{Vis}$–*wet* and bamp$_{Aud}$damp$_{Vis}$–*middle* in Block 1, and bamp$_{Aud}$bamp$_{Vis}$–*middle* and bamp$_{Aud}$damp$_{Vis}$–*wet* in Block 2; another subject received the same stimuli in the reverse order. Each participant received 192 trials: half McGurk primes and half Congruent primes, and, orthogonally, half word targets (half each related and unrelated to the prime) and half nonword targets. (Note that the nonword targets were also presented following critical prime stimuli – which primes were paired with word versus nonword targets was counterbalanced across subjects.) Order of presentation of stimuli within each block was randomized for each subject. The experiment began with seven practice trials.

2.3. Results

Lexical decision reaction times (RTs) were measured from the offset of the target item. Only RTs following critical (i.e., real-word target) stimuli were analyzed. Responses that were incorrect (3.7%), occurred before target onset (1.5%), were over two standard deviations from the within-subject, within-category mean (3.9%), or were mis-recorded due to equipment error (3.1%) were excluded. Three items with missing data were excluded in the items analysis.

Remaining RTs were submitted to 2 (Relatedness: Related, Unrelated) × 2 (Prime Congruency: Congruent, McGurk) × 2 (Prime Auditory Lexicality: auditory-nonword, auditory-word) ANOVAs by subjects (F_1) and items (F_2). Participant means are shown in Fig. 2.

As expected, a priming effect was observed: Related targets (M = 178 ms) were responded to faster than Unrelated targets (M = 256 ms; main effect of Relatedness: $F_1(1, 23) = 74.807$, $p < .0001$; $F_2(1, 43) = 18.508$, $p < .0001$). Responses to targets following McGurk-prime stimuli (dotted lines in Fig. 2; M = 210 ms) trended toward faster responses than targets following Congruent-prime stimuli (solid lines in Fig. 2; M = 224 ms), though only by subjects (main effect of Prime Congruency: $F_1(1, 23) = 4.223$, $p = .051$; $F_2(1, 43) < 1$). Target responses following auditory-word stimuli (right side of Fig. 2; M = 209 ms) were faster than those following auditory-nonword stimuli (left side of Fig. 2; M = 226 ms) (main effect of Prime Auditory Lexicality by subjects: $F_1(1, 23) = 7.522$, $p = .012$; but not by items: $F_2(1, 43) < 1$). There was a significant Prime Congruency × Prime Auditory Lexicality interaction by items ($F_2(1, 43) = 6.558$, $p = .014$), though not by subjects ($F_1(1, 23) = 2.380$, $p = .137$). None of the remaining interactions were significant by either subjects or items (all Fs < 1.046; all ps > .317).

Given this interaction and the predicted differences between the two groups of McGurk stimuli, planned, separate 2 (Relatedness) × 2 (Prime Congruency) ANOVAs were conducted by subjects and by items for auditory-nonword and auditory-word prime stimuli to examine the effects of each type of Prime Auditory Lexicality individually.

For auditory-nonword primes (left side of Fig. 2), responses to Related targets (M = 191 ms) were faster than Unrelated targets (M = 260 ms; $F_1(1, 23) = 32.905$, $p < .0001$; $F_2(1, 21) = 8.418$,

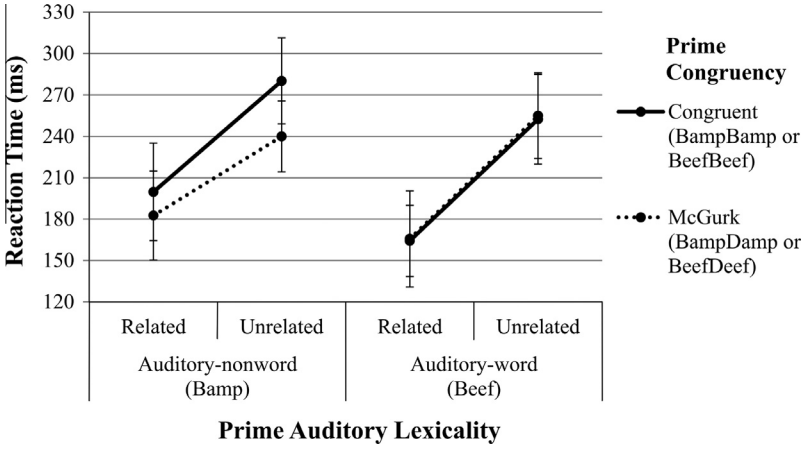

Fig. 2. Lexical decision reaction times on target words, as a function of Prime Congruency, Prime Auditory Lexicality, and Relatedness. The priming effect for each condition is represented by the slope of the line. Error bars show standard error of the mean.

p = .009). Targets following McGurk items (perceived as real words, e.g., $bamp_{Aud}damp_{Vis}$ → $damp_{Percept}$; M = 211 ms) were responded to more quickly than targets following Congruent items (which were nonwords, e.g., $bamp_{Aud}bamp_{Vis}$ → $bamp_{Percept}$; M = 240 ms; $F_1(1, 23) = 6.197$, $p = .021$; $F_2(1, 21) = 7.400$, $p = .013$). Targets following McGurk primes and targets following Congruent primes produced the same priming effect (no Relatedness × Prime Congruency interaction: both Fs < 1).

For auditory-word primes (right side of Fig. 2), responses to Related targets (M = 165 ms) were faster than Unrelated targets (M = 253 ms; $F_1(1, 23) = 40.295$, $p < .0001$; $F_2(1, 22) = 10.233$, $p = .004$). There was no difference between targets following McGurk or Congruent primes and no interaction (all Fs < 1.2). Critically, in this group, McGurk and Congruent primes shared identical, real word auditory signals.

2.4. Discussion

Experiment 1 had two main findings. For auditory-word primes (e.g., $beef_{Aud}beef_{Vis}$ and $beef_{Aud}deef_{Vis}$), prime congruency did not affect lexical decision RTs: Targets following McGurk items (auditory words perceived as nonwords) and Congruent items (auditory words perceived as words) were responded to equally quickly. As the McGurk and Congruent items had the identical auditory signals but different visual signals, this pattern of results means that a stimulus's visual signal did not affect lexical access to the primes and RTs to the targets. This suggests that the *auditory signal* is the lookup key in the lexicon and AV-integration occurs post-lexically.

However, for auditory-nonword primes (e.g., $bamp_{Aud}bamp_{Vis}$ and $bamp_{Aud}damp_{Vis}$), targets following McGurk items (auditory nonwords perceived as words) were responded to more quickly than targets following Congruent items (auditory nonwords perceived as nonwords). Again, the auditory signals did not differ between corresponding McGurk and Congruent items, but for these stimuli, the differing visual signals did affect RTs and thus lexical access. This response facilitation for stimuli that integrate to form words suggests that the *integrated McGurk stimulus* is used for lexical access and thus AV-integration occurs pre-lexically. As participants were equally susceptible to the McGurk Effect for both types of McGurk stimuli (given equivalent ratings in the pilot experiment), the discrepancy cannot be explained by subjects integrating the auditory and visual signals for one group but not the other.

Taken together, these data suggest that the relative time course of AV-integration and lexical access depends on the lexicality of the auditory signal, such that AV-integration and lexical access occur in parallel and compete in a race to completion. That is, AV-integration takes some time to complete, and while it is progressing, lexical search begins using the auditory signal due to its privileged status in speech comprehension. If the auditory signal represents a word and thus finds a match in the lexicon, that word and its semantic associates are activated, and lexical access is complete. AV-integration, however, is not, and although a word has already been accessed, integration continues and ultimately produces the fused result the listener consciously perceives.

However, if the auditory signal is a nonword, lexical access based on the auditory signal cannot succeed on its own. In this case, the lexical search can be updated with new information from the integrated McGurk percept (once AV-integration completes) to help disambiguate the unclear auditory signal.

Due to the extremely short inter-stimulus interval between prime and target in the present experiment (50 ms), the longer processing time required for AV-integration could easily extend into the time of target presentation. As a result, the relative speed of access of the prime could affect RTs to closely-following targets.

This process can also account for the main effect of Prime Auditory Lexicality. Auditory-nonword McGurk stimuli ($bamp_{Aud}damp_{Vis}$) must wait for AV-integration to complete before lexical access occurs, and thus should produce slower RTs to subsequent target items than auditory-word McGurk stimuli ($beef_{Aud}deef_{Vis}$), for which lexical access completes before AV-integration does. This matches the pattern found in Experiment 1.

One facet of the data that deserves mention is the lack of Prime Congruency × Relatedness interaction for the auditory-nonword primes: That is, Congruent ($bamp_{Aud}bamp_{Vis}$) and McGurk ($bamp_{Aud}damp_{Vis}$) items produced an equivalent degree of priming (even if the Congruent stimuli led to slower target reaction times overall). Why do these Congruent stimuli work as primes, since they are both heard and consciously perceived as nonwords? This is likely due to the robust effect of mediated phonological-to-semantic priming. The signature of mediated priming is that nonwords that differ from real words by one phonetic feature (like the Congruent auditory-nonword stimuli such as *bamp* in the present experiment, which differ from the real words like *damp* only by place of articulation) elicit slower overall responses but still significant priming effects (Connine, Blasko, & Titone, 1993; Marslen-Wilson, Moss, & van Halen, 1996; Milberg, Blumstein, & Dworetzky, 1988). That is, reaction times are faster to a one-feature-away nonword followed by a related target (e.g., *pomato-ketchup*) than to a pair of unrelated words (e.g., *shoe-ketchup*), thus demonstrating a priming effect. However, although the one-feature-away nonwords still elicit significant priming, they do produce numerically slower reaction times overall than do prime-target pairs consisting of two related real words (e.g., *tomato-ketchup*). This corresponds to the pattern observed in the present experiment's results: *wet* is numerically slower following $bamp_{Aud}bamp_{Vis}$ than $bamp_{Aud}damp_{Vis}$, but both types of primes facilitate responses to the related (*wet*) over the unrelated (*middle*) targets. This explains why $bamp_{Aud}bamp_{Vis}$ works as a prime at all.

Given that the nonword primes produce mediated priming, how can we be sure that the critical observed differences (that is, the different behavior following auditory-word [$beef_{Aud}deef_{Vis}$] and auditory-nonword [$bamp_{Aud}damp_{Vis}$] McGurk primes) are not also due to mediated priming? If the differences between the McGurk and Congruent primes were attributable to mediated priming from the auditory signal (namely, that hearing *bamp* or *damp* is equivalent), we would expect no difference between responses to targets presented after $bamp_{Aud}bamp_{Vis}$ and after $bamp_{Aud}damp_{Vis}$, as they have the identical nonword auditory track (and thus the identical mediated priming effect). However, this is not the case: RTs are significantly slower following $bamp_{Aud}bamp_{Vis}$ (cf. the main effect of Prime Congruency in the auditory-nonword sub-ANOVA), which we explain as having resulted from the RT slowdown observed to one-feature-away nonwords. Similarly, if the critical effects were attributable to mediated priming from the consciously perceived (McGurk) signal, we would expect no difference between responses to related targets presented after $bamp_{Aud}bamp_{Vis}$ and after $beef_{Aud}deef_{Vis}$, as both are perceived as one-feature-away nonwords. Again, this is not the case: RTs are slower following $bamp_{Aud}bamp_{Vis}$ than following $beef_{Aud}deef_{Vis}$, numerically (200 ms and 166 ms, respectively) and at marginal significance ($t(23) = 1.84$, $p = .078$).

The results from Experiment 1 suggest that listeners can perceive one word (the McGurk percept) while lexically accessing another (the auditory signal). Experiment 2 tests this hypothesis using stimuli for which both the auditory signal and McGurk percept were real words. A semantic priming paradigm used McGurk and Congruent primes, and targets related to either the McGurk auditory signal or percept. We predict that although subjects will consciously perceive the fused McGurk stimulus, the auditory signal's lexical representation will be accessed. Therefore, we expect

that McGurk and Congruent primes with identical auditory signals will produce the same pattern of priming.

3. Experiment 2

Experiment 2 tested the prediction that a McGurk item with an auditory-word signal activates that word's lexical entry, regardless of the listener's ultimately-reported perception. If the relative time course of AV-integration depends on the lexicality of the auditory stream, then McGurk stimuli with auditory-word signals should prime targets semantically related to the auditory signal, even while subjects report perceiving something different: namely, the fused McGurk percept.

3.1. Method

3.1.1. Participants

One hundred forty-four[2] students at the University of California, San Diego (49 male; 17–33 years old) participated. One subject was replaced due to missing data in one condition. All were native English speakers.

3.1.2. Materials

Stimuli consisted of an auditory–visual prime followed, after a 50 ms ISI, by an auditory-only target. McGurk primes were composed of auditory and visual signals differing only on the initial consonant, and all auditory and visual signals and combined McGurk percepts were common English words. As in Experiment 1, the auditory and visual signals differed only by the onset consonant's place of articulation, and thus the intended McGurk percept matched the visual signal for each item.

Items that successfully induced a McGurk percept were selected from a pilot experiment. Percept "goodness" scores were calculated by averaging ratings of the acceptability of the McGurk percept with the *unacceptability* of the auditory signal. Thus, goodness scores measured the relative acceptability of the McGurk percept and auditory signal, and were high for items perceived as McGurk fusions and *not* as the auditory signals. Auditory goodness scores were calculated similarly.[3]

Thirty-six McGurk stimuli were selected on the basis of these ratings (e.g., $bait_{Aud}date_{Vis} \rightarrow date_{Percept}$) from the 118 McGurk stimuli presented in the pilot experiment. Each McGurk prime had two congruent AV counterparts: one matching the auditory signal ("CongA": $bait_{Aud}bait_{Vis} \rightarrow bait_{Percept}$) and one matching the visual signal ("CongV": $date_{Aud}date_{Vis} \rightarrow date_{Percept}$). For McGurk items, percept goodness scores were significantly higher ($M = 3.84/5$) than auditory goodness scores ($M = 2.16/5$; $t(35) = 13.73$, $p < .001$), demonstrating that the selected stimuli successfully induced a McGurk Effect. Additionally, the fusion onset consonant was rated higher than the auditory onset consonant for all selected McGurk items. The scores for the two associated congruent items were not different (CongA: $M = 4.63/5$, CongV: $M = 4.72/5$, $t(35) = 1.07$, $p = .293$).

Semantically related target words were selected for both auditory and visual components of the McGurk primes (and thus corresponding congruent primes) using the same databases as in Experiment 1, and an online norming study of UCSD students ($N = 151$) for primes not available in either database. Thirty-six nonword targets from the same nonword database as was used in Experiment 1 were paired with non-critical McGurk and corresponding congruent primes to create filler items for the lexical decision task. All nonwords were pronounceable and one or two syllables.

Each critical auditory–visual prime was matched with four auditory-only word targets: one related to the auditory track of the associated McGurk stimulus, one related to the visual track of the associated McGurk stimulus, and two unrelated targets. The two unrelated targets for each prime were assigned by shuffling the Auditory-Related targets and re-pairing them to different prime items to become Auditory-Unrelated targets, and shuffling the Visual-Related targets and re-pairing them to become Visual-Unrelated targets. Thus each target item was its own control: *worm* appeared as an Audio-Related target for the McGurk prime $bait_{Aud}date_{Vis} \rightarrow date_{Percept}$ and (for a different subject) as an Audio-Unrelated target for the McGurk prime $mine_{Aud}nine_{Vis} \rightarrow nine_{Percept}$. Each item family therefore consisted of 12 prime-target pairs: 3 prime items (McGurk, CongA, CongV) crossed with 4 target items (Auditory-Related, Auditory-Unrelated, Visual-Related, Visual-Unrelated); see Fig. 3, and Appendix B for a complete list of stimuli.

3.2. Procedure

The experimental procedure was identical to that of Experiment 1. The design differed in that participants never received the same prime or target twice; additionally, if they saw a McGurk prime, they did not see either congruent counterpart. Half the subjects heard a given target as a related word; the other half, as an unrelated word. Subjects received three stimuli in each condition; thus, a total of 36 critical primes paired with real-word targets. All subjects received the same 36 nonword-target filler trials, which were equally divided between McGurk, CongA, and CongV primes.

Subjects were instructed to watch the speaker's face and perform a lexical decision as quickly as possible on the target. The experiment began with six practice trials with feedback.

3.3. Results

Lexical decision RTs were measured from the offset of the target item. Only RTs following critical (real-word target) stimuli were analyzed. Responses that were incorrect (7.5%), occurred before target onset (0.1%), or were more than two standard deviations from their condition means (3.9%) were excluded.[4] Five subjects (in the F_1 analysis) and one item (in the F_2 analysis) which were

[2] Each prime-target pair belonged to an item family comprising 12 conditions, and each subject saw only one item from each group. Therefore, there were only three observations per condition from each subject, and thus many more subjects were required than for Experiment 1.

[3] Note that this is a slightly modified (and we believe, improved) stimulus selection criterion from the one used for Experiment 1. To ensure that the results of Experiment 1 still held when applying this more apt selection procedure, all Experiment 1 analyses (F1 and F2 omnibus and sub-ANOVAs) were recalculated using only the subset of items that would have been included had the Experiment 2 selection criterion been in place during Experiment 1. This better selection procedure was applied in two ways and the Exp 1 results recalculated for each: first, by excluding stimuli for which the visual (McGurk) rating (Exp 1 criterion) differed from the percept goodness score (Exp 2 criterion) by more than 1 point, and second, by excluding all items for which the composite percept goodness score (Exp 2 criterion) was *lower* than the raw visual rating score (Exp 1 criterion). For this post hoc procedure, the idea was to exclude any items that demonstrated less good AV-integration as measured by the better rating system (goodness score) than it appeared to (raw visual rating) for inclusion in Exp 1. Under the first exclusion criterion, all main effects and interactions in all F1 and F2 analyses produced the identical level of significance (significant, marginal, or not significant) as reported in Exp 1. Under the second exclusion criterion, just two effects changed significance level: the interaction between Prime Congruency x Prime Auditory Lexicality in the F1 omnibus ANOVA changed from not significant to significant ($F_1(1, 23) = 5.072$, $p = .034$), and the main effect of Prime Auditory Lexicality in the F2 omnibus ANOVA changed from not significant to marginal ($F_2(1, 27) = 3.211$, $p = .084$).

[4] As noted, each subject was presented with just three trials in each condition. Therefore, the preferred criterion for excluding outliers – based on the within-subject, within-condition means (as was done for Exp 1) – was impractical for Exp 2. However, recalculating the results of Exp 1 using the same outlier criterion as Exp 2 (namely, excluding RTs more than two standard deviations from their within-condition mean) again resulted in the identical level of significance (significant, marginal, or not significant) for all F1 analyses, both the omnibus and sub-ANOVAs.

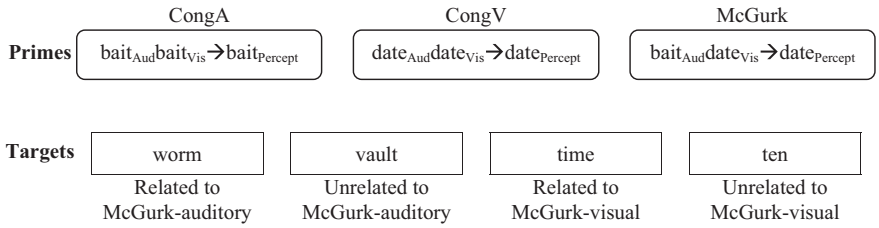

Fig. 3. Stimuli from one item family. All primes in an item family were paired with all targets in that item family, resulting in 12 prime-target pairs (3 primes crossed with 4 targets).

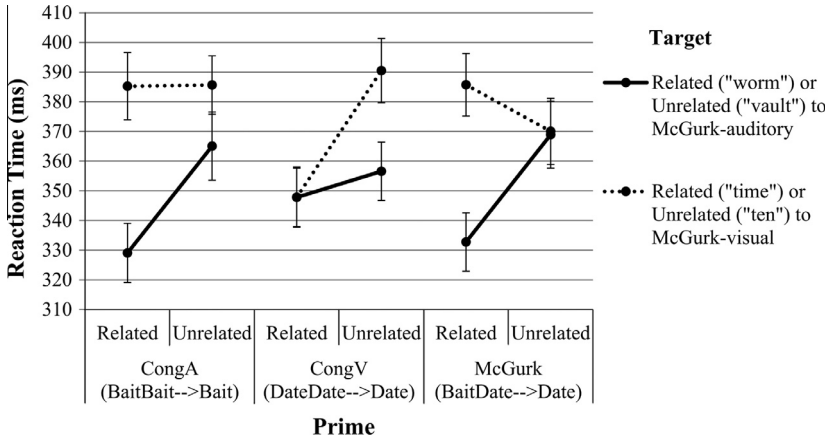

Fig. 4. Lexical decision reaction times on target words, as a function of Prime, Target, and Relatedness. The priming effect for each condition is represented by the slope of the line. Error bars show standard error of the mean.

missing observations in at least one cell were excluded. The remaining critical RTs were submitted to 2 (Relatedness: Related, Unrelated) × 3 (Prime: McGurk, CongA, CongV) × 2 (Target: McGurk-auditory, McGurk-visual) omnibus ANOVAs by subjects (F_1) and items (F_2). Participant means are shown in Fig. 4.

Again, as expected, there was a strong priming effect (Related: $M = 355$ ms, Unrelated: $M = 373$ ms; main effect of Relatedness: $F_1(1, 138) = 12.088$, $p < .0007$; $F_2(1, 69) = 20.570$, $p < .0001$). Collapsed across Target and Relatedness conditions, subjects responded with the same latency regardless of prime type (CongA: $M = 366$ ms, CongV: $M = 361$ ms, McGurk: 364 ms; no main effect of Prime: both Fs < 1). Subjects responded faster to targets associated with the auditory component of the McGurk stimulus (solid lines in Fig. 4; $M = 350$ ms) than with the visual component (dotted lines in Fig. 4; $M = 378$ ms; main effect of Target by subjects: $F_1(1, 138) = 45.550$, $p < .0001$; but not by items: $F_2(1, 69) = 2.668$, $p < .107$).

Targets associated with the McGurk auditory signal (both related: *worm* and unrelated: *vault*; solid lines in Fig. 4) elicited comparable RTs regardless of prime type (CongA: $M = 347$ ms; CongV: $M = 352$ ms; McGurk: $M = 351$ ms), whereas targets associated with the McGurk visual signal (*time* and *ten*; dotted lines in Fig. 4) elicited faster RTs following CongV primes ($M = 369$ ms) than following CongA primes ($M = 385$ ms) or McGurk primes ($M = 378$ ms), though this difference was only significant when calculated by items and not by subjects (Prime × Target interaction: $F_1(2, 276) = 1.539$, $p = .216$; $F_2(2, 138) = 3.987$, $p = .021$).

The three prime conditions elicited the same degree of priming, as facilitation for Related over Unrelated targets did not differ following CongA primes (18 ms), CongV primes (26 ms), or McGurk primes (10 ms; no Prime × Relatedness interaction: both Fs < 1.2; both ps > .309).

There was a larger priming effect for targets related to the McGurk auditory signal (Related [*worm*]: $M = 337$ ms, Unrelated [*vault*]: $M = 364$ ms) than the McGurk visual signal (Related [*time*]: $M = 373$ ms, Unrelated [*ten*]: $M = 382$ ms; Target × Relatedness interaction: $F_1(1, 138) = 3.307$, $p = .071$; $F_2(1, 69) = 5.828$, $p = .018$). This interaction arises because two-thirds of the stimuli produced a larger priming effect for the McGurk-auditory targets (CongA and McGurk primes; left and right portions of Fig. 4) and only one-third produced a larger priming effect for the McGurk-visual targets (CongV primes; center portion of Fig. 4). Therefore, when averaged across all three prime types, the priming effect for the targets associated with the McGurk-auditory signal should indeed be larger than that for the targets associated with the McGurk-visual signal.

The critical, predicted priming effect was observed through a significant three-way interaction between Relatedness, Prime, and Target ($F_1(2, 276) = 10.147$, $p < .0001$; $F_2(2, 138) = 6.656$, $p = .002$). Following CongA primes (bait$_{Aud}$bait$_{Vis}$ → bait$_{Percept}$), targets that were Related versus Unrelated to the McGurk-auditory track showed *more* facilitation (*worm-vault*: 36 ms priming) than did targets that were Related versus Unrelated to the McGurk-visual track (*time-ten*: 0 ms priming). Following CongV primes (date$_{Aud}$date$_{Vis}$ → date$_{Percept}$), targets that were Related versus Unrelated to the McGurk-auditory track showed *less* facilitation (*worm-vault*: 9 ms priming) than did targets that were Related versus Unrelated to the McGurk-visual track (*time-ten*: 43 ms priming). These two results demonstrate that the targets selected to be related to the McGurk-auditory and the McGurk-visual were indeed well-chosen to elicit selective effects of semantic priming from those respective signals. Crucially, McGurk primes (bait$_{Aud}$date$_{Vis}$ → date$_{Percept}$) matched the pattern of CongA primes: targets that were Related versus Unrelated to the McGurk-auditory

Table 2
F_1 effects for prime-pairwise Relatedness × Prime × Target comparisons.

	McGurk vs. CongA primes		CongA vs. CongV primes		McGurk vs. CongV primes	
Effect	F_1	p	F_1	p	F_1	p
Relatedness[a]	**5.651**	**.019**	**13.440**	**.0004**	**9.448**	**.003**
Prime[b]	0.147	.702	1.327	.251	0.474	.492
Target[c]	**34.924**	**<.0001**	**25.340**	**<.0001**[†]	**18.582**	**<.0001**[†]
Relatedness × Prime	0.606	.438	0.545	.462	2.488	.117
Relatedness × Target	**12.919**	**.0005**	0.005	.946	0.910	.342
Prime × Target	0.906	.343	3.834	.052[*]	0.547	.461[*]
Relatedness × Prime × Target	0.562	.455	**12.069**	**.0007**	**20.809**	**<.0001**

Note. Degrees of freedom for all F_1 cells are (1, 138). Significant effects are in bold. All F_2 effects showed comparable significance, except those indicated: Effects marked with daggers were significant in F_1 but not F_2 analyses; effects marked with asterisks were significant in F_2 but not F_1 analyses. Note that "Target" is a within-subjects comparison but a between-items comparison.

[†] CongA vs. CongV: Target: $F_2(1, 69) = 2.172$, $p = .145$; McGurk vs. CongV: Target: $F_2(1, 69) = 1.602$, $p = .210$.

[*] CongA vs. CongV: Prime × Target: $F_2(1, 69) = 6.785$, $p = .011$. McGurk vs. CongV: Prime × Target: $F_2(1, 69) = 4.366$, $p = .040$.

[a] Related vs. Unrelated.

[b] Column headers.

[c] McGurk-auditory vs. McGurk-visual.

track showed *more* facilitation (*worm-vault*: 36 ms priming) than did targets related versus unrelated to the McGurk-visual track (*time-ten*: −16 ms priming).

The critical result hinges upon whether the priming effect produced by McGurk primes matches that of CongA or CongV primes – namely, whether the priming behavior, and thus lexical access, of the McGurk primes is driven by the auditory signal or the McGurk percept. To directly compare effects from each prime type, planned 2 (Relatedness) × 2 (Prime) × 2 (Target) ANOVAs on RTs for each pair of prime conditions were conducted by subjects (F_1) and by items (F_2). Results from the three F_1 ANOVAs are presented in Table 2.

As demonstrated in Fig. 4 and Table 2, McGurk primes patterned with CongA primes: for both, targets related to the McGurk auditory track (solid lines in Fig. 4) were primed more than targets related to the McGurk visual track (dotted lines in Fig. 4). The reverse was true for CongV primes: targets related to the McGurk visual track were primed more than targets related to the McGurk auditory track. These contrasting patterns gave rise to a significant Relatedness × Target interaction when comparing McGurk and CongA primes, demonstrating that for both of these prime conditions, targets associated with the McGurk-auditory track produced a substantial priming effect, whereas targets associated with the McGurk-visual track did not. This interaction was not significant when comparing CongV primes to CongA primes, or CongV primes to McGurk primes, because both targets related to the McGurk-auditory and to the McGurk-visual tracks, when collapsed across these two prime condition pairs, produced a priming effect from their respective prime condition.

Most tellingly, the Relatedness × Prime × Target interactions were significant when comparing CongV primes to either CongA or McGurk primes, but there were no such differences when comparing CongA and McGurk primes to each other. Thus, McGurk and CongA primes elicited the same priming behavior as each other, but CongV and CongA primes, and CongV and McGurk primes, did not elicit the same priming behavior.

3.4. Discussion

Congruent primes matching the McGurk auditory signal ($bait_{Aud}bait_{Vis} \rightarrow bait_{Percept}$) produced greater priming for targets related to the auditory signal (*worm*), and congruent primes matching the McGurk visual signal ($date_{Aud}date_{Vis} \rightarrow date_{Percept}$) produced greater priming for targets related to the visual signal (*time*). This demonstrates a basic semantic priming effect. Critically, McGurk primes ($bait_{Aud}date_{Vis} \rightarrow date_{Percept}$) matched the pattern of CongA primes, showing a large priming effect for targets related to the auditory signal (*worm*) and no priming for targets related to the McGurk percept and visual signal (*time*). This suggests that lexical access for McGurk stimuli is driven by the auditory signal.

These data support those from Experiment 1 and suggest that one word (corresponding to the McGurk auditory signal) may be lexically accessed, while another (corresponding to the fused McGurk percept) is consciously perceived. Following McGurk primes, facilitation was observed for targets semantically related to the McGurk auditory signals but not the McGurk visual signals/percepts, suggesting that with auditory-word stimuli, only the auditory signal accesses the lexicon. However, the two perceptual streams eventually integrate, producing the perceptual experience of the McGurk illusion.

4. General discussion

In Experiment 1, Congruent and McGurk primes with real word auditory signals elicited equivalent target reaction times. Conversely, primes with nonword auditory signals elicited target RTs that were slower overall, and dissociated such that primes that were ultimately perceived as real words elicited faster responses than those perceived as nonwords. In Experiment 2, McGurk stimuli primed semantic associates of their auditory-word signals, despite the fact that these items were perceived as different words that were created from integrating the mismatching auditory and visual streams.

When does AV-integration occur relative to lexical access? The answer appears to be somewhat more complicated than a strictly early or late process; that is, always before or always after lexical access has occurred. The priming results from Experiment 2 support a post-lexical integration account – priming was observed for semantic associates of the auditory signals (but not of the McGurk fusions); thus lexical access must have occurred on the auditory signal alone and AV-integration must have completed *after* lexical access. Similarly, the auditory-word stimuli in Experiment 1 showed that the identity of the visual signal did not affect the priming of related target words (prime: $beef_{Aud}deef_{Vis} \rightarrow deef_{Percept}$; target: *meat*), suggesting that lexical accessed occurred before the visual signal was integrated with the auditory. However, the differential priming patterns elicited by Congruent and McGurk

auditory-nonword stimuli in Experiment 1 showed that the identity of the visual signal *did* affect the priming of related target words (prime: $bamp_{Aud}damp_{Vis} \rightarrow damp_{Percept}$; target: *wet*), suggesting that lexical access occurred *after* the visual signal was integrated with the auditory. Thus there is evidence for both pre- and post-lexical AV-integration within the present paradigm (and within prior literature as well).

What, then, determines the relative time course of AV-integration and lexical access? When a comprehender is presented with a multimodal AV speech signal, the auditory and visual signals are necessarily processed separately, as they come in through different modalities and neural streams. Lexical access begins, using the more informative auditory signal. AV-integration begins as well, but is slower to complete.

When the auditory signal has a match in the lexicon – namely, it is a real word – that word is selected, activating its semantically-related associates. If one of those associates must be responded to shortly thereafter (as in the present experiments in which the target followed the prime after only 50 ms), AV-integration has not yet had a chance to influence the lexical access process by the time of the response decision. AV-integration eventually completes, producing the McGurk illusion. If a word has already been accessed in the lexicon (corresponding to the auditory signal), then AV-integration only affects the listener's eventual conscious perception, because there is no longer pending lexical access activity to resolve. Thus, although the item corresponding to the McGurk illusion is consciously perceived, it is either not lexically accessed at all, or to such a minimal degree that it does not sufficiently activate its semantic network to generate priming (at least during the initial process that a 50 ms ISI taps into). This explains the finding from Experiment 1 that stimuli with auditory-word signals are responded to identically regardless of their ultimately-perceived lexicality ($beef_{Aud}deef_{Vis} \rightarrow deef_{Percept}$ and $beef_{Aud}beef_{Vis} \rightarrow beef_{Percept}$), and the finding from Experiment 2 that McGurk stimuli ($bait_{Aud}date_{Vis} \rightarrow date_{Percept}$) prime targets related to their auditory signals (*worm*) but not their visual signals (*time*). This similarly is supported by prior evidence that when the auditory and visual signals are not presented simultaneously, people predominantly perceive the auditory signal when it is presented earlier than the visual signal, and predominantly perceive a McGurk Effect fusion when the auditory signal is delayed relative to the visual signal (e.g., Munhall, Gribble, Sacco, & Ward, 1996; van Wassenhove, Grant, & Poeppel, 2007). When presentation of the auditory signal is delayed, so is lexical access, giving sufficient time for the visual signal to be processed.

However, if a word has not yet been accessed by the time AV-integration completes – for example, when the auditory signal is a nonword – then the integrated percept can provide additional information and thus affect lexical access. This explains the result from Experiment 1 that targets following primes with nonword auditory signals ($bamp_{Aud}damp_{Vis} \rightarrow damp_{Percept}$) were responded to more slowly than targets following primes with word auditory signals ($beef_{Aud}deef_{Vis} \rightarrow deef_{Percept}$), as the former need AV-integration to complete before lexical access can succeed. It additionally explains the result from Experiment 2 that, even with minimal priming of targets related to the visual signal (*time*), subjects' conscious perception of McGurk stimuli was influenced by this stream. Thus, these experiments demonstrate that AV-integration and lexical access are interdependent, and their relative time course (namely, whether lexical access occurs before or after AV-integration) depends on the lexical properties of the stimulus.

We should note that it is possible, especially in light of the effects of mediated priming (as discussed above), that the word corresponding to the visual signal (e.g., "date") is accessed along with the auditory signal (e.g., "bait") during the first-pass lexical access, though not sufficiently to elicit semantic priming. As lexical access is generally viewed as a competitive process (e.g., Marslen-Wilson, 1987, and many others), regardless of whether the visual signal is activated minimally or not at all, the activation of the word corresponding to the auditory signal is greater than that of the visual signal and wins the competition.

These results, of course, hinge upon subjects actually perceiving the McGurk Effect: perhaps subjects never integrated the mismatching auditory and visual signals of the McGurk stimuli, and auditory-word-related targets were primed because subjects perceived only the auditory stream. It is impossible to guarantee that every subject perceived the illusion on every trial. However, it is unlikely that participants as a whole were not susceptible to the effect, as only items that reliably elicited a McGurk Effect in each respective pilot experiment (and for which the McGurk percept was rated higher than the auditory signal) were included as stimuli in the main experiments. Furthermore, there was a difference in responses following McGurk primes and Congruent primes with the same auditory track in the condition where the fused percept was expected to influence reaction times – namely, in the auditory-nonword condition in Experiment 1 ($bamp_{Aud}damp_{Vis}$ and $bamp_{Aud}bamp_{Vis}$ primes). If participants did not perceive the McGurk Effect, then the McGurk and Congruent trials in this condition should have produced the same response times (given that they had the same auditory signals), contrary to what was observed.

Previous evidence supports the claim that visual information can play a role during delayed lexical access. Fort et al. (2013) showed participants visual-only (i.e., mouthed) primes which matched or mismatched the first syllable of high- or low-frequency target words. Matching visual stimuli facilitated low- but not high-frequency targets. The authors suggest that because lexical access takes longer for low-frequency target words, the visual-only primes selectively influenced these items. These primes are an extreme version of the current study's auditory-nonword primes: Both have poor auditory signals which may delay lexical access of the prime, thereby giving the visual signal enough time to influence this lexical access process.

Similarly, Brancazio (2004) divided his participants' reaction times to report their perception of McGurk stimuli into slow, medium, and fast responses. For fast responses, there was no effect of the lexicality of either the auditory or visual signal – regardless of whether the stimulus fused into a word or nonword, participants were equally likely to report a visual response. However, for medium and slow responses – when the visual signal had more time to be integrated with the auditory signal – participants' perceptual reports were more susceptible to lexical bias, and they were more likely to report they had perceived the stimulus's visual signal when it was a word compared to a nonword. The present account predicts that AV integration has a larger effect at longer latencies in terms of lexical access; Brancazio (2004) finds this in terms of reported percepts.

Other studies purporting to show the influence of visual information on lexical access provide contradictory evidence. Some (e.g., Barutchu et al., 2008; Brancazio, 2004) found that a mouthed word influences perception of an auditory-nonword input, and is more likely to be integrated if the auditory signal is "bad" and integration produces perception of a real word, while others have not (e.g., Sams et al., 1998). Similarly, Bastien-Toniazzo, Stroumza, and Cavé (2009) found higher incidence of McGurk percepts in noisy environments, implying that less reliable auditory signals cause greater reliance on visual signals. These studies, however, assessed perceptual identification of stimuli rather than examining whether they accessed the

lexical-semantic network. Our results suggesting that stimuli with different auditory and visual lexicalities can produce conflicting lexical access and perception may provide the link to reconcile these inconsistent findings.

Converging evidence is also provided by a recent paper by Samuel and Lieblich (2014), which supports the dissociation between lexical access and perception for auditory–visual stimuli. The researchers used a selective adaptation paradigm, in which exposure to an adaptor syllable changes the way listeners identify other test syllables. In two experiments, participants provided a baseline categorization of test stimuli that fell along an auditory continuum (e.g., classifying items between /bI/ and /dI/ as either "B" or "D"). Then, participants were presented with the adaptors – auditory–visual stimuli which elicited illusory percepts (e.g., arma**b**illo$_{Aud}$arma**g**illo$_{Vis}$). The participants' perceptual reports for the illusory AV stimuli ("arma**d**illo") were different from the raw auditory signal ("arma**b**illo"). Participants categorized the test stimuli throughout the experiment, providing a measure of adaptation relative to their baseline. Critically, both the illusory percepts (arma**b**illo$_{Aud}$arma**g**illo$_{Vis}$) and the corresponding auditory-only stimuli (arma**b**illo$_{Aud}$Ø$_{Vis}$) failed to induce phonological adaptation. In contrast, normally-produced (Congruent) words that matched the illusory percept (e.g., arma**d**illo$_{Aud}$arma**d**illo$_{Vis}$) *did* elicit adaptation. This means that *perception* of "arma**d**illo" did not itself cause adaptation, even with the lexical context provided by AV fusion; rather, selective adaptation was determined by the identity of a stimulus's auditory signal. These data thus support the present work suggesting that the auditory signal can be used for lexical processing, while the combined auditory–visual signal is ultimately consciously perceived. Roberts and Summerfield (1981) found similar results using nonword adaptors, and showed that an illusory AV stimulus elicits selective adaptation matching the pattern elicited by the auditory signal, but not the perceived signal (cf. Saldaña & Rosenblum, 1994).

Thus, the evidence points to a dissociation between lexical access and conscious perception in situations in which the former can complete before the latter has time to do so. A distinction of this nature is not without precedent; for example, there have been prior demonstrations of multiple words being accessed in parallel, irrespective of the ultimate conscious percept. For example, hearing a word that contains another word as one of its syllables (such as "trombone", which includes "bone") primes associates of the contained word (e.g., "dog"), even though the conscious percept is that of the entire containing word (e.g., Vroomen & de Gelder, 1997). Embedded words like "bone" not only activate their semantic associates in the lexicon, but are also (temporarily) integrated into the ongoing sentential context (van Alphen & van Berkum, 2010). Similar to the present research, these results suggest that listeners are "updating possible interpretations using all available information as soon as they can" (van Alphen & van Berkum, 2010, p. 2625) – in this latter case by maintaining multiple lexical options ("bone", "trombone"), and in the present research, by maintaining multiple sensory options (auditory "bait", visual "date"). However, there is evidence that onset embeddings, such as "pill" in "pilgrim," can be disambiguated by acoustic or sentential cues and thus although multiple words are initially activated simultaneously, activation for the shorter, embedded word quickly falls off (Davis, Marslen-Wilson, & Gaskell, 2002; Salverda, Dahan, & McQueen, 2003). Findings of this type demonstrate that multiple words can be lexically accessed, at least briefly, during spoken word recognition, when those words actually are embedded within the auditory input signal.

The multiple activation demonstrated in these prior studies is essentially the reverse of the results presented here. In the case of words like "trombone," a single input stimulus initially accesses multiple competing lexical items ("trombone," "bone," "own") which eventually get pruned to a single activated word (presumably the target word "trombone.") However, in the present research, multiple input stimuli (auditory "bait," visual "date") initially access a single lexical item ("bait"). Perhaps eventually, after AV-integration completes, the perceived word ("date") is accessed as well. Thus, prior work on multiple activation has shown that a single input stimulus can initially access multiple lexical items and later only a single word is activated. The present work shows that two competing input stimuli can initially access a single lexical item and later, perhaps, a second word is additionally activated.

Similarly, in standard cohort competitor effects, an onset syllable that is shared by multiple words (e.g., "cap", shared by "captain" and "captive") produces on-line activation of the cohort competitors (as measured by eye movements, e.g., Allopenna, Magnuson, & Tanenhaus, 1998), and also primes associates of both ("ship" and "guard"; e.g., Marslen-Wilson, 1987; Yee & Sedivy, 2006). Both cohort competitors appear to be activated during the initial stages of word recognition, even when presented in a sentence context that biases toward one word but not the other (Zwitserlood, 1989). Thus, not only can multiple words be accessed from the same auditory signal when they are contained by the sound form ("bone" and "trombone"), but also as long as the sound form is consistent with those multiple words ("cap" is the onset to both "captain" and "captive").

These prior studies demonstrate that a *single input stimulus* can activate multiple other words in parallel, some of which are accessed or recognized but do not reach the level of conscious perception. The present result, while related, is importantly distinct as it demonstrates that in the case where there are *multiple input stimuli* that are conflicting and mutually exclusive, one of those stimuli may reach conscious perception, whereas the other may reach the level of word recognition in the absence of conscious perception. In the case of cohort competitors, although multiple words are indeed activated and lexically accessed in parallel (as evidenced by the eye-tracking and semantic priming results), it is not the case that one of the competitors is lexically accessed while the other is consciously perceived. Rather, a third (shared) stimulus – "cap" in this example – causes both competitors to be accessed due to their phonological overlap. The present work shows a dissociation between the fates of two competing input streams (e.g., hearing "bamp" and seeing "damp"); prior work has shown that a single stimulus can cause perceivers to activate multiple related words (if you heard "trombone", you necessarily perceived "bone" as well, and "cap" is indeed contained in both "captain" and "captive"). Similarly, the present work addresses the process of combining and resolving information from multiple sources to enable word recognition. The research on multiple activation and cohort competitor effects investigates the process of selecting the correct word from a field of competitors as the input unfolds.

Auditory–visual integration occurs in parallel with lexical access. Integration may or may not affect which word is ultimately accessed: sometimes the accessed word is the same as the perceived word; sometimes it is not. Thus, listeners can lexically access one word but consciously perceive another.

Acknowledgements

Thanks to Elena Tenenbaum and Naomi Feldman for guidance, Liz Chrastil and Clarice Robenalt for speech modeling, Matt Wynn and Vicky Tu for videography, John Mertus for technical skills, and the *CL research assistants for testing subjects. This work was supported in part by NIH Grants DC00314 to S.E.B. and HD32005 to J.L.M.

Appendix A. Stimuli used in Experiment 1

McGurk primes were formed by combining the auditory recording of "Prime Auditory" with the visual recording of "Prime Visual." Targets were presented auditory-only.

Prime auditory	Prime visual	Semantically related target	Semantically unrelated target
beef	deef	meat	ask
bamp	damp	wet	middle
bance	dance	music	name
bawn	dawn	morning	name
bense	dense	thick	off
best	dest	worst	chap
busk	dusk	night	off
mall	nall	shop	doing
meck	neck	head	own
mend	nend	fix	down
miece	niece	nephew	ought
mife	(k)nife	fork	middle
milk	nilk	cow	eight
mind	nind	brain	end
mist	nist	fog	fact
month	nonth	year	find
most	nost	all	fine
moun	noun	verb	point
mow	now	later	own
much	nuch	lot	hand
mug	nug	beer	ought
munch	nunch	eat	heard
mute	nute	deaf	her
pag	tag	label	put
pain	tain	hurt	here
pal	tal	friend	left
pame	tame	wild	point
parp	tarp	tent	round
pask	task	job	put
pass	tass	fail	let
path	tath	trail	line
paunt	taunt	tease	saw
pave	tave	road	little
paw	taw	dog	live
pay	tay	money	live
peam	team	ball	round
peen	teen	age	saying
peeth	teeth	gums	saw
pext	text	book	seem
pierce	tierce	ear	long
pig	tig	hog	long
pight	tight	loose	saying
pime	time	clock	should
pink	tink	blue	means
pongue	tongue	mouth	seem
poss	toss	throw	side
powel	towel	dry	should
pump	tump	gas	means

Appendix B. Stimuli used in Experiment 2

McGurk primes were formed by combining the auditory recording of "Prime Auditory" with the visual recording of "Prime Visual." CongA primes were congruent items formed by combining the auditory and visual recordings of "Prime Auditory" words. CongV primes were congruent items formed by combining the auditory and visual recordings of "Prime Visual" words. Targets were presented auditory-only.

Prime auditory	Prime visual	Auditory-related target	Auditory-unrelated target	Visual-related target	Visual-unrelated target
dad	bad	mom	nothing	good	wind
bait	date	worm	vault	time	ten
dank	bank	dark	stir	money	no
bay	day	water	pan	night	tiny
bead	deed	necklace	dog	will	lymph
dean	bean	school	mom	green	good
beer	deer	drink	paper	doe	blood
dell	bell	wood	kill	ring	person
bet	debt	gamble	road	owe	hammer
bid	did	auction	pea	done	joint
bill	dill	payment	rat	pickle	stomach
pad	tad	paper	drill	little	doe
part	tart	piece	preen	sour	exam
pie	tie	apple	dark	neck	money
pole	toll	vault	yours	booth	time
pot	tot	pan	crowd	tiny	door
pug	tug	dog	method	pull	will
pest	test	bug	piece	exam	hair
mail	nail	letter	gamble	hammer	sleep
maim	name	kill	hobo	person	chew
map	nap	road	letter	sleep	owe
me	knee	you	auction	joint	fish
mice	nice	rat	cigarette	mean	pickle
night	might	day	sell	maybe	girl
nil	mill	nothing	school	wind	green
mine	nine	yours	worm	ten	booth
mix	nix	stir	apple	no	neck
mob	nob	crowd	water	door	night
mode	node	method	necklace	lymph	pull
primp	crimp	preen	bug	hair	sour
bore	gore	drill	drink	blood	little
bum	gum	hobo	wood	chew	ring
bun	gun	bread	day	bullet	maybe
pod	cod	pea	you	fish	done
butt	gut	cigarette	payment	stomach	mean
buy	guy	sell	bread	girl	bullet

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38*(4), 419–439.

Baart, M., & Samuel, A. G. (2015). Turning a blind eye to the lexicon: ERPs show no cross-talk between lip-read and lexical context during speech sound processing. *Journal of Memory and Language, 85*, 42–59.

Barutchu, A., Crewther, S. G., Kiely, P., Murphy, M. J., & Crewther, D. P. (2008). When /b/ill with /g/ill becomes /d/ill: Evidence for a lexical effect in audiovisual speech perception. *European Journal of Cognitive Psychology, 20*(1), 1–11.

Bastien-Toniazzo, M., Stroumza, A., & Cavé, C. (2009). Audio-Visual perception and integration in developmental dyslexia: An exploratory study using the McGurk effect. *Current Psychology Letters, 25*(3) (Online).

Besle, J., Fort, A., Delpuech, C., & Giard, M.-H. (2004). Bimodal speech: Early suppressive visual effects in human auditory cortex. *European Journal of Neuroscience, 20*(8), 2225–2234.

Braida, L. D. (1991). Crossmodal integration in the identification of consonant segments. *The Quarterly Journal of Experimental Psychology, 43*(3), 647–677.

Brancazio, L. (2004). Lexical influences in audiovisual speech perception. *Journal of Experimental Psychology. Human Perception and Performance, 30*(3), 445–463.

Colin, C., Radeau, M., Soquet, A., & Deltenre, P. (2004). Generalization of the generation of an MMN by illusory McGurk percepts: Voiceless consonants. *Clinical Neurophysiology, 115*(9), 1989–2000.

Colin, C., Radeau, M., Soquet, A., Demolin, D., Colin, F., & Deltenre, P. (2002). Mismatch negativity evoked by the McGurk–MacDonald effect: A phonetic

representation within short-term memory. *Clinical Neurophysiology, 113*(4), 495–506.
Connine, C. M., Blasko, D. G., & Titone, D. (1993). Do the beginnings of spoken words have a special status in auditory word recognition? *Journal of Memory and Language, 32*(2), 193–210.
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.
Erber, N. P. (1975). Auditory–visual perception of speech. *Journal of Speech and Hearing Disorders, 40*(4), 481–492.
Fort, M., Kandel, S., Chipot, J., Savariaux, C., Granjon, L., & Spinelli, E. (2013). Seeing the initial articulatory gestures of a word triggers lexical access. *Language and Cognitive Processes, 28*(8), 1207–1223.
Fort, M., Spinelli, E., Savariaux, C., & Kandel, S. (2010). The word superiority effect in audiovisual speech perception. *Speech Communication, 52*(6), 525–532.
Grant, K. W., & Seitz, P. F. (2000). The use of visible speech cues for improving auditory detection of spoken sentences. *The Journal of the Acoustical Society of America, 108*(3 Pt 1), 1197–1208.
Green, K. P. (1998). The use of auditory and visual information during phonetic processing: Implications for theories of speech perception. In R. Campbell, B. Dodd, & D. Burnham (Eds.), *Hearing by Eye II: Advances in the psychology of speechreading and auditory–visual speech* (pp. 3–26). Hove, England: Psychology Press.
Kiss, G. R., Armstrong, C., Milroy, R., & Piper, J. (1973). An associative thesaurus of English and its computer analysis. In A. J. Aitken, R. W. Bailey, & N. Hamilton-Smith (Eds.), *The Computer and Literary Studies*. Edinburgh: University Press.
Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition, 25*(1–2), 71–102.
Marslen-Wilson, W., Moss, H. E., & van Halen, S. (1996). Perceptual distance and competition in lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 22*(6), 1376–1392.
Massaro, D. W. (1987). *Speech perception by ear and eye: A paradigm for psychological inquiry*.
Massaro, D. W., & Jesse, A. (2007). Audiovisual speech perception and word recognition. In M. G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (pp. 19–35). New York: Oxford University Press.
McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*(5588), 746–748.
Milberg, W., Blumstein, S., & Dworetzky, B. (1988). Phonological factors in lexical access: Evidence from an auditory lexical decision task. *Bulletin of the Psychonomic Society, 26*(4), 305–308.
Miller, G. A., & Nicely, P. E. (1955). An analysis of perceptual confusions among some English consonants. *The Journal of the Acoustical Society of America, 27*(2), 338–352.
Munhall, K. G., Gribble, P., Sacco, L., & Ward, M. (1996). Temporal constraints on the McGurk effect. *Perception & Psychophysics, 58*(3), 351–362.
Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). *The University of South Florida word association, rhyme, and word fragment norms*. <http://www.usf.edu/FreeAssociation/>.
Rastle, K., Harrington, J., & Coltheart, M. (2002). 358,534 nonwords: The ARC nonword database. *The Quarterly Journal of Experimental Psychology Section A, 55*(4), 1339–1362.
Roberts, M., & Summerfield, Q. (1981). Audiovisual presentation demonstrates that selective adaptation in speech perception is purely auditory. *Perception & Psychophysics, 30*(4), 309–314.
Saint-Amour, D., De Sanctis, P., Molholm, S., Ritter, W., & Foxe, J. J. (2007). Seeing voices: High-density electrical mapping and source-analysis of the multisensory mismatch negativity evoked during the McGurk illusion. *Neuropsychologia, 45*(3), 587–597.
Saldaña, H. M., & Rosenblum, L. D. (1994). Selective adaptation in speech perception using a compelling audiovisual adaptor. *Journal of the Acoustical Society of America, 95*(6), 3658–3661.
Salverda, A. P., Dahan, D., & McQueen, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition, 90*(1), 51–89.
Sams, M., Manninen, P., Surakka, V., Helin, P., & Kättö, R. (1998). McGurk effect in finnish syllables, isolated words, and words in sentences: Effects of word meaning and sentence context. *Speech Communication, 26*(1–2), 75–87.
Samuel, A. G., & Lieblich, J. (2014). Visual speech acts differently than lexical context in supporting speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 40*(4), 1479–1490.
Soto-Faraco, S., Navarra, J., & Alsius, A. (2004). Assessing automaticity in audiovisual speech integration: Evidence from the speeded classification task. *Cognition, 92*(3), B13–B23.
Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *The Journal of the Acoustical Society of America, 26*(2), 212–215.
Summerfield, Q. (1987). Some preliminaries to a comprehensive account of audio-visual speech perception. In B. Dodd & R. Campbell (Eds.), *Hearing by Eye: The psychology of lip-reading* (pp. 3–51). Hillsdale, NJ: Lawrence Erlbaum Associates Ltd..
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.
van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory–visual speech perception. *Neuropsychologia, 45*, 598–607.
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.
Windmann, S. (2004). Effects of sentence context and expectation on the McGurk illusion. *Journal of Memory and Language, 50*(2), 212–230.
Yee, E., & Sedivy, J. C. (2006). Eye movements to pictures reveal transient semantic activation during spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(1), 1–14.
Zwitserlood, P. (1989). The locus of the effects of sentential-semantic context in spoken-word processing. *Cognition, 32*(1), 25–64.