Published in final edited form as:
Lang Cogn Process. 2009 ; 24(4): 580–610. doi:10.1080/01690960802536357.

Visual speech primes open-set recognition of spoken words

Adam B. Buchwald,
Department of Speech-Language Pathology and Audiology, New York University, New York, NY, and Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN, USA

Stephen J. Winters, and
Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN, USA, and Department of Linguistics, University of Calgary, Calgary, AB, Canada

David B. Pisoni
Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN, USA

Abstract

Visual speech perception has become a topic of considerable interest to speech researchers. Previous research has demonstrated that perceivers neurally encode and use speech information from the visual modality, and this information has been found to facilitate spoken word recognition in tasks such as lexical decision (Kim, Davis, & Krins, 2004). In this paper, we used a cross-modality repetition priming paradigm with visual speech lexical primes and auditory lexical targets to explore the nature of this priming effect. First, we report that participants identified spoken words mixed with noise more accurately when the words were preceded by a visual speech prime of the same word compared with a control condition. Second, analyses of the responses indicated that both correct and incorrect responses were constrained by the visual speech information in the prime. These complementary results suggest that the visual speech primes have an effect on lexical access by increasing the likelihood that words with certain phonetic properties are selected. Third, we found that the cross-modality repetition priming effect was maintained even when visual and auditory signals came from different speakers, and thus different instances of the same lexical item. We discuss implications of these results for current theories of speech perception.

Keywords

Visual speech; Audiovisual priming; Word recognition; Lexical access

INTRODUCTION

Over the past two decades, characterisations of speech perception and the sensory processing of speech signals have become more focused on the multimodal nature of speech

Correspondence should be addressed to Adam B. Buchwald, Department of Speech-Language Pathology and Audiology, New York University, 665 Broadway, Suite 910, New York, NY 10012, USA. buchwald@nyu.edu.

Publisher's Disclaimer: This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.
The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

(Bernstein, 2005; Calvert, Spence, & Stein, 2004; Kim, Davis, & Krins, 2004; Massaro, 1987, 1998; Massaro & Cohen, 1995; Massaro & Stork, 1998; Rosenblum, 2005; Summerfield, 1987). One line of work on this topic has explored the processing of linguistic information conveyed in visual speech, including both the ability to identify words from visual-only speech signals (e.g., Auer & Bernstein, 1997; Auer, 2002; Lachs, Weiss, & Pisoni, 2000; Mattys, Bernstein, & Auer, 2002) as well as the ability to identify the language being spoken (for adults: Soto-Faraco et al., 2007; Ronquest, Levi, & Pisoni, 2007; for infants: Weikum, Vouloumanos, Navarra, Soto-Faraco, Sebastián-Gallés, & Werker, 2007). Research in this vein has revealed that the linguistic information conveyed in visual-only speech signals facilitates subsequent processing of the same words presented in the auditory modality (Kim et al., 2004; also see Dodd, Oerlemens, & Robinson, 1989) suggesting that cross-modality priming effects in speech perception may occur at modality-independent levels of representation, and providing additional support for the claim that auditory and visual speech are processed using a common recognition system (Auer, 2002; Mattys et al., 2002; Rosenblum, 2005; Rosenblum, Miller, & Sanchez, 2007).

In this paper, we report investigations on how visual speech information and auditory speech information are encoded and processed using a variant of a short-term priming paradigm (e.g., Foster & Davis, 1984). Priming studies are typically used in psycholinguistics to address issues of whether and when certain representations are active in the course of language processing (e.g., McLennan, Luce, & Charles-Luce, 2003; also see papers in Bowers & Marsolek, 2003 for more discussion). Typically, researchers examine changes in responses to a ‘target’ stimulus when the target is preceded by a ‘prime’ stimulus, and these changes (typically in response time or response accuracy) reflect the relationship between the target and prime stimuli in the cognitive processing required for the task. In the present work, we employed a cross-modality repetition priming paradigm with visual speech lexical primes and auditory speech lexical targets, similar to Kim et al. (2004) who measured differences in lexical decision performance. Following their work, if visual speech perception and auditory speech perception rely on shared cognitive and neural resources (e.g., the same amodal representations at some level of processing), we should observe facilitatory effects of the visual speech prime on recognition of the auditory speech target.

We extended the Kim et al. (2004) results by examining participants' performance in an open-set spoken word recognition task, comparing trials when a noise-degraded auditory word is preceded by a silent video clip of the same word to a baseline condition in which the auditory target is preceded by a still image of the speaker. Using an open-set word recognition task permitted us to compare both the rate of correct responses and the properties of incorrect responses. Each of these comparisons would indicate whether a priming effect of visual speech on auditory word recognition occurs at a modality-independent level of representation. While this claim is supported by the work reported in Kim et al. (2004) and reviewed below, it remains unclear whether the critical level of lexical representation is part of an instance-based (or exemplar) lexicon, or is at a somewhat more abstract (i.e., removed from the signal) level of representation. In a second experiment, we asked whether this type of cross-modality priming occurs at a level of episodic representation by investigating whether a necessary condition for the cross-modality priming effect was that the visual and auditory signals were coming from the same speech ‘episode’. If the data reveal that the priming effect is not attenuated when the visual and auditory speech come from different sources, then this would indicate that the priming effect occurs at a level of representation that is neither modality-specific nor episodic.

Visual speech primes and auditory speech targets

Previous research has demonstrated effects of visual speech primes on auditory targets using a variety of speech perception tasks. Dodd, Oerlemans, and Robinson (1989) observed lexical repetition priming effects with visual-only primes and auditory-only targets. Using a semantic categorisation task, Dodd et al. reported faster reaction times to auditory lexical targets when participants had previously been presented with a block of visual-only lexical primes compared with a control condition with no block of primes. This finding suggests that the visual prime and the auditory target activate common semantic representations in memory. It is worth noting here that Dodd et al. used visual speech stimuli which were readily identified on their own by at least 80% of participants in a screening task, indicating that the priming effect could have come from separate identification of each of the two stimuli.

More recently, Kim et al. (2004) had participants perform a lexical decision task on spoken words that were preceded by visual-only speech signals which were not readily identifiable in isolation. Kim et al. compared participants' reaction times in a lexical decision task on trials with a lexically consistent visual speech prime (i.e., the same lexical item) to trials with lexically inconsistent visual speech primes; they reported facilitation (i.e., faster reaction times) in the responses for trials with words (but not nonwords) with consistent visual speech primes. They concluded that speech perception is ‘amodal’ because the priming effect suggests that visual and auditory signals activate common representations. It is possible that the difference between the baseline and experimental conditions in Kim et al.'s study was due to response inhibition in the presence of inconsistent stimulus information as opposed to response facilitation in the presence of consistent information. Nevertheless, both of these explanations suggest the presence of some type of common representation activated by both the visual and auditory signals which affects processes used to perform the lexical decision task.

The present paper reports on two experiments. The first experiment was performed as a replication of Kim et al.'s study using an open-set word recognition task rather than a lexical decision task. The use of the open-set word recognition task allowed us to investigate the level at which the priming occurs, as we can determine whether all responses (right and wrong) are influenced by the visual speech prime. If the visual speech prime affects lexical access by activating all words that share phonological features with the visual speech prime, then we would expect the target word to be easier to identify given this ‘pre’-activation. In addition, we would expect that even incorrect responses will be closer to the target with respect to sub-phonemic features available in visual speech, as the participants should be more likely to respond with a word that shares the targets visual speech features (or visemes). Expanding on the notion of the perceptual equivalence class from Miller and Nicely (1955; also see Huttenlocher & Zue, 1984; Shipman & Zue, 1982), Auer and Bernstein (1997) developed the construct of lexical equivalence class (also see Lachs et al., 2000; Mattys et al., 2002), which is an equivalence class for words that are indistinguishable from the visual speech stream (e.g., *pin* and *bin* which differ only in voicing, a feature that is not detectable in visual speech). We will use this notion to help assess the hypothesis that the visual speech primes affect lexical access by activating all words that are consistent with the visual-only speech clip.

Episodic accounts of lexical access in speech perception

The nature of the mental lexicon has been the subject of intense scrutiny and spirited debate in psycholinguistics since Oldfield (1966) described this construct as a collection of words in long-term memory that mediates access between perception and lexical knowledge. Restricting our discussion to the sound structure representations in the mental lexicon, three main views (broadly speaking) have emerged in recent years regarding lexical representation: the *abstractionist* view, in which words in the lexicon are represented as being composed of linear sequences of context-free units (e.g., Morton, 1979); the *episodic* (or exemplar) view, in which lexical representation consists of an encoding of detailed memory traces of each word (e.g., Goldinger, 1998; Johnson, 2005); and a *hybrid* or *mixed* view, in which both instance-specific and abstract, general information is encoded (e.g., Pierrehumbert, 2001; see Pisoni & Levi, in press for a review of several proposals).

Goldinger (1996, 1998; Goldinger & Azuma, 2003) has argued extensively for the claim that perceivers store instance-specific exemplars of linguistic input that they encounter and that these exemplars form the lexicon (also see Bybee, 2001; Johnson, 1997, 2005; Pierrehumbert, 2001; Port & Leary, 2005; Sheffert & Fowler, 1995 among others). In this view, an individual word does not have a single representation, but is associated with an exemplar ‘cloud’ consisting of detailed memory traces of the episodic experiences the language user has had with that word. This account is consistent with performance on a variety of language processing tasks in which speakers' perception and production of a word change over the course of more exposure to (and use of) that word. Larger changes have been observed for low frequency words which are assumed to have less robust representations due to the lower number of previously encountered tokens of those words.

Although Goldinger's work (among others) suggests that we do encode the particular experiences we encounter and that these episodes form part of our lexical knowledge, there exists other evidence that lexical access also involves some representations of linguistic structure that are abstractions over the encountered exemplars. In a recent example using repetition priming, McLennan et al. (2003) reported that intervocalic /t/ produced both carefully and as its casual speech allophone – flap – prime one another in a repetition priming experiment, suggesting that they both activate a shared mediating, abstract (i.e., non instance-specific) representation of /t/.

Given evidence that both exemplars and abstractions over those exemplars are stored in the lexicon, we are interested in determining the level at which the cross-modality priming effect discussed here and by Kim et al. (2004) arises. While Kim et al.'s data suggested that the priming effect occurs at a level that is not modality-specific, it remains possible that this amodal level of representation is still instance-based. For example, both auditory and visual speech signals could be translated into amodal representations prior to storage in an exemplar-based lexicon. Under this scenario, it is possible that the priming effect is observed *because* the visual and auditory speech signals come from the same real world event, and are effectively stored as identical traces in exemplar memory. If this is true, event identity of auditory and visual speech tokens is not merely a sufficient condition, but is a necessary condition to generate the repetition priming effect. The plausibility of this account is suggested by the recent work of Lachs and Pisoni (2004a, 2004b) who demonstrated that observers can match a visual speech display to one of two voices in ABX or XAB tasks, thus indicating that we are able to encode speaker-specific information from one modality and reliably map it onto the same speaker in a different modality (we will address this further in the general discussion).

EXPERIMENT 1

In Experiment 1, we sought to replicate the cross-modality priming in speech perception findings of Kim et al. (2004) using a task that allows us to explore the nature of the priming effect at a finer grain of analysis. We employed an open-set spoken word recognition task with auditory targets preceded by either visual speech primes of the same speaker producing the same word or by a still picture of the talker's face. One critical goal of this experiment – beyond that reported by Kim et al. (2004) – was to gain insight into the nature of the priming effect not only by examining differences in overall spoken word recognition accuracy, but also investigating whether the responses on trials with visual speech primes are more constrained than responses on trials in the control condition, and whether the nature of these constraints is predictable by (and can shed light on) the linguistic information present in visual speech signals. If this information constrains responses by activating the forms that are consistent with the visual speech prime, we should observe both higher overall accuracy (a categorical measure) and incorrect responses that are closer to the target with respect to sublexical features (a gradient measure).

In addition, the stimuli we used in this study were selected to test whether the improvement in spoken word recognition performance with the visual prime interacts with lexical factors (i.e., lexical frequency, neighbourhood density) that are known to influence the efficiency and accuracy of word recognition for both auditory speech (e.g., Luce & Pisoni, 1998) and visual speech (e.g., Auer, 2002). In particular, it has been demonstrated that ‘easy’ words (i.e., high frequency words from sparse neighbourhoods) are recognised quickly and more accurately than ‘hard’ words (i.e., low frequency words from dense neighbourhoods) in both auditory stimuli (e.g., Luce & Pisoni, 1998) and in visual speech perception (e.g., Auer, 2002). If the present study obtained a stronger cross-modality priming effect for ‘easy’ words than for hard words, then this would support the claim that the repetition priming effect occurs at a level of cognitive processing that interacts with lexical access.

Method

Participants—Forty Indiana University undergraduate students, ages 18–23, participated in Experiment 1. All participants were native speakers of English with no speech or hearing disorders. Participants received either course credit or monetary compensation for their participation in this study.

Materials—All stimulus materials were drawn from the Hoosier multi-talker audio-visual (AV) database (Sheffert, Lachs, & Hernandez, 1997). Monosyllabic, CVC words produced by one female speaker and one male speaker in the database were selected for this study. The stimulus set for each participant contained 96 different word tokens (see Appendix for stimulus list). In each condition, half of the stimuli were ‘Easy’ words – high frequency words from lexically sparse phonological neighbourhoods (e.g., ‘fool’), while the other half were ‘Hard’ words – low frequency lexical items from lexically dense phonological neighbourhoods (e.g., ‘hag’; see Luce & Pisoni, 1998).

Auditory stimuli: In each condition, we used envelope-shaped noise (Horii, House, & Hughes, 1971) to reduce performance on the spoken word recognition task. The experimental stimuli were created by processing the audio files through a MATLAB script that randomly changed the sign bit of the amplitude level of 30% of the spectral samples in the acoustic waveform. Reducing auditory-only word recognition performance to below-ceiling levels is a necessary prerequisite to detect the effects of cross-modality repetition priming in the spoken word recognition task. Pilot data indicated that this level of noise degradation reduced auditory-only open-set recognition to about 50% correct.

Visual stimuli: Two kinds of visual primes were used: visual speech and control. Visual speech primes consisted of the original, unedited video clips associated with each target word. Previous research has shown that the overall identification accuracy on these stimuli presented in a visual-only condition was 14%, with less than 1% of the individual tokens accurately identified more than 90% of the time (Lachs & Hernandez, 1998). Thus, the specific words used in the study could not be consistently identified in a visual-only condition, although it is worth noting that these analyses were only performed with respect to words correct, and does not take into account the issue of lexical equivalence classes (e.g., Auer & Bernstein, 1997). The video track of the control primes consisted of a still shot of the speaker whose duration was identical to that of its counterpart in the visual speech prime condition. The same still image was used in the control condition for each target word. This image was taken from a resting state of each speaker. The still image was used as a control rather than a video clip of a different word as our pilot testing indicated that participants largely ignored the video clip primes when some were lexically consistent with the target and others were lexically inconsistent with the target.[1]

Procedure—Participants were tested in groups of four or fewer in a quiet room with individual testing booths. During testing, each participant listened to the auditory signals over Beyer Dynamic DT-100 headphones at a comfortable listening level while sitting in front of a Power Mac G4. A customised SuperCard (v4.1.1) stack presented the stimuli to each participant. Participants were instructed to watch the computer monitor and then type the English word that they heard over the headphones using the computer keyboard.

On each trial (see Figure 1), participants first saw either a visual speech prime or the control prime. Five hundred milliseconds after the presentation of the visual prime, participants heard the degraded auditory target word over the headphones. This inter-stimulus interval (ISI) was used following the procedure of Lachs and Pisoni (2004a, 2004b). A prompt then appeared on the screen asking the participant to type the word they heard. The participant's responses were recorded on a keyboard, and the presence of a response on the keyboard was measured 60 times per second, yielding a possible error of +/− 16 ms for reaction times; these data were not subjected to further analysis. Presentation of the next stimulus was participant-controlled.

Participants were either presented with all female talker stimuli (both targets and primes) or all male talker stimuli. Words were presented to participants in random order, with Dynamic and Control primes randomly interleaved over the course of the experiment. Each participant responded to 48 words in each priming condition, half of which were lexically ‘Easy’ targets and half of which were lexically ‘Hard’ targets.

Results

Word identification accuracy—For analyses reported in this section, the dependent variable was spoken word recognition accuracy. The results revealed that the participants benefited from the presentation of the Visual speech prime when compared with the Control prime. Overall, participants in Experiments 1 exhibited a 14% accuracy gain on trials in which the Visual speech prime preceded the degraded audio signal (67%) compared with trials in the control condition (53%). The word recognition accuracy data for the female and male talkers were analysed with separate 2 × 2 Prime type (Visual speech/Control) vs. Target type (Easy/Hard) repeated measures Analyses of Variance (ANOVAs). The ANOVAs revealed a significant main effect of Prime type for both the female speaker: Visual speech = 66.8% (*SD* = 9.3%), Control = 49.1% (*SD* = 9.8%); $F_1(1, 19) = 166.8$, $p < .001$; $F_2(1, 46) = 38.6$, $p < .001$; and the male speaker: Visual speech = 67.2% (*SD* = 11.1%); Control = 56.7% (*SD* = 6.8%); $F_1(1, 19) = 166.7$, $p < .001$; $F_2(1, 46) = 22.9$, $p < .001$, as well as significant main effects of Target type for both speakers – Female speaker: Easy = 67.1% (*SD* = 14.4), Hard = 49.9% (*SD* = 13.8); $F_1(1, 19) = 121.2$, $p < .001$, $F_2(1, 46) = 43.2$, $p < .001$; Male speaker: Easy = 69.9% (*SD* = 12.6%), Hard = 52.5% (*SD* = 11.4%), $F_1(1, 19) = 196.7$, $p < .001$; $F_2(1, 46) = 38.9$, $p < .001$. Thus, better performance was obtained on trials with Visual speech primes compared with trials with Control primes, and

[1] This was observed in pilot data for participants who saw trials with lexically consistent Visual Primes, lexically inconsistent dynamic speech primes and control static images. For these participants, there were no performance differences among these three groups. In addition, several pilot participants reported that they were not attending to the prime after realising it was sometimes not the target word. The inconsistent stimuli in the pilot were other words from the study that were randomly selected for display; thus, each word was CVC, but no other factors with respect to similarity were controlled.

on trials with Easy targets compared with trials with Hard targets. The interaction between Prime type and Target type was not significant for either speaker when analysed by subject or by items.

Response analysis: Experiment 1—To enrich our understanding of the information observers perceive and encode in the Visual speech prime condition when compared with the Control prime condition, we performed several analyses comparing the responses participants made on Visual speech trials to those made on Control trials. For the purposes of increasing power over the analyses, the data from participants who observed the Female speaker and those who observed the Male speaker were combined for all analyses reported in this section.

Collapsing over all the data, there were 1920 responses for each trial type. A total of 465 unique responses[2] were given for Visual speech trials, whereas 610 unique responses were given for Control trials. A chi-square analysis revealed that significantly more unique responses were provided in response to Control trials than to Visual speech trials, $\chi^2(1) = 46.40$, $p < .01$. This finding strengthens the word identification accuracy results reported above and indicates that the information present in the Visual speech prime constrains the participants' responses to the auditory word presented in noise. Additionally, we observed significantly fewer unique responses on trials with Easy targets (476) compared with trials with Hard targets: 599; $\chi^2(1) = 19.23$, $p < .01$. The difference in the number of unique responses for Easy words with the two prime types (Visual speech: 190; Control: 286), and Hard words with the two prime types (Visual speech: 275; Control: 324) approached but did not reach significance, $\chi^2(1) = 3.68$, $p < .06$.

Additional analyses were designed to explore the nature of the constraints on the response selection process. Initially, each response was coded for the number of correct segments of the CVC word (i.e., 0–3 segments correct), and the average number of correct segments for each participant was then computed for each condition. This provided a more detailed measure of overall response accuracy. A repeated measures ANOVA with Prime type (Visual speech vs. Control) and Target type (Easy vs. Hard) as independent variables and overall segmental accuracy as the dependent variable revealed a main effect for Prime type, $F(1, 39) = 75.81$, $p < .001$, with higher segmental accuracy observed for targets with Visual speech primes (mean = 2.49, $SD = 0.20$) than for targets with Control primes (mean = 2.21, $SD = 0.19$). The ANOVA also revealed a main effect of Target type, $F(1, 39) = 69.46$, $p < .001$, with significantly higher segmental accuracy observed for Easy targets (mean = 2.44, $SD = 0.28$) than for Hard targets (mean = 2.25, $SD = 0.24$). There was also a significant interaction between Prime type and Target type, $F(1, 39) = 12.57$, $p < .001$. The locus of the interaction indicated that the effect of Target type was larger for the Visual speech primes (Easy: 2.62; Hard: 2.34) than for the Control primes (Easy: 2.26; Hard: 2.16). Overall, these results suggest that the responses on trials with Visual speech primes were more constrained (i.e., closer to the target) than the trials in the control condition.

To further address whether incorrect responses were also more constrained when preceded by Visual speech primes, we limited the analysis described above to responses in which the participant gave the wrong whole word response (thus giving a possible range of 0–2 segments correct). Using the number of correct segments in incorrect responses as the dependent variable, we performed 2 × 2 Prime Type (Visual speech/Control) vs. Target type

[2] Unique responses are specific target-response combinations. For example, the target word 'watch' was produced accurately in the control condition by all subjects except for one, who wrote 'what'. Thus, we considered this to be two unique responses to the word 'watch' – 'watch' and 'what'. If another subject responded with 'what' to the target stimulus of 'wash', we called this an additional unique response as this is a different response-target pair.

(Easy/Hard) repeated measures ANOVA. This analysis also revealed a significant main effect of Prime type, $F(1, 39) = 15.35$, $p < .001$; the number of correct segments on trials with Visual speech primes (mean = 1.47, $SD = 0.18$) was significantly greater than the number of correct segments on trials with Control primes (mean = 1.34, $SD = 0.14$). This result indicates that the information present in the Visual speech video signal constrains all of the participants' responses, leading to greater accuracy even for incorrect responses.

Additionally, a main effect of Target type was obtained, $F(1, 39) = 23.56$, $p < .001$; however, when the analysis was limited to incorrect responses, the responses to trials with Hard targets had significantly higher overall segmental accuracy (mean = 1.49, $SD = 0.21$) than responses on trials with Easy targets (mean = 1.29, $SD = 0.30$). This result may at first appear surprising; however, it reflects a significantly higher proportion of incorrect responses with 2 segments correct on trials with Hard targets, 558/927, 60.2%, than Easy targets, 257/621, 41.4%; $\chi^2(1) = 52.02$, $p < .001$. Given the current operational definition of lexical neighbours as words sharing N − 1 segments of an N-segment word (which was used to generate the Easy/Hard targets for this experiment; see Luce & Pisoni, 1998), participants' incorrect responses that contained two correct segments are, by definition, lexical neighbours of the target word. Thus, incorrect responses to Hard targets (words from lexically dense neighbourhoods) were more likely to be neighbours of the target than incorrect responses on trials with Easy targets (words from lexically sparse neighbourhoods). The interaction between Prime type and Target type was significant, $F(1, 39) = 6.78$, $p < .05$, with the effect of Target type attenuated for Visual speech prime trials (Easy: 1.43; Hard: 1.53) compared with Control prime trials (Easy: 1.16; Hard: 1.46). Thus, the effect of Target type was stronger in the condition when there was no Visual speech i.e., visual information about the target, further suggesting that this additional optical information provides a constraint on participants' open-set word identification responses.

The above results reveal that incorrect responses on trials with Visual speech primes are closer to the target than incorrect responses on trials with Control primes. To gain a more detailed understanding of how the Visual speech information constrains responses on the word recognition task, we analysed the likelihood of correct responses for each syllable position of the CVC words as a function of Prime type. These data, presented in Table 1, show that the accuracy is greater for words in the Visual Speech condition compared with the Control condition for each of the syllable positions, revealing that the Visual speech information helped constrain responses for all three syllabic positions of the CVC words.

To determine whether there was a difference in the accuracy gain for any of the three positions, we computed a difference score (Visual speech–Control) for each syllable position. Planned comparisons indicated that the cross-modality priming effect was significantly greater for onset position than it was for either nucleus position, $t(39) = 2.51$, $p < .05$, or for coda position, $t(39) = 3.58$, $p < .01$, but there was no difference between accuracy on the nucleus position and coda position, $t(39) = 1.49$, *ns*.

The data analysed in this section thus far suggest that there was a global benefit from the Visual speech primes which constrained all components of the participants' responses, and that this effect was particularly robust for onset position. However, it remains possible that the information in the Visual speech prime constrained responses by limiting specific components of the set of competing hypotheses about the target word. To address this possibility, we examined the participants' identification of specific phono-logical properties of the target stimulus. Specifically, we examined the likelihood that participants would correctly identify the place features (divided into three categories: *labial*, *coronal*, and *dorsal*), manner features (divided into five categories: *stop*, *fricative*, *nasal*, *liquid*, and *glide*), and voicing features (divided into two categories: *voiced* and *voiceless*) of the onset and coda consonants in the target word. These analyses were performed by collapsing the data obtained from all 40 subjects, and comparing the accuracy on these individual dimensions for target words with Visual speech primes and target words with Control primes. The results of these analyses are presented in Table 2.

The data in Table 2 indicate that the Visual speech primes created a robust increase in accuracy with respect to place and manner of articulation for both onset and coda consonants. This result suggests that the participants were able to use the optical information available in the Visual speech prime to limit the set of possible responses, and that this information was useful in specifying both place and manner of articulation. With respect to voicing, we limited our analysis to those trials in which the target and response were obstruents and thus the voice feature would have to be specified as part of the response. Although we failed to observe a significant effect of prime type on accuracy of voicing, it is worth noting that performance on this feature is near ceiling even in the control condition.

These results were further augmented by an analysis that examined the likelihood that incorrect responses were part of the same lexical equivalence class as the target (Auer & Bernstein, 1997; Lachs et al., 2000; Mattys et al., 2002). A paired *t*-test indicated that trials with Visual speech primes were more likely to have incorrect responses that were part of the same lexical equivalence class as the target, mean = 28.7%, *SD* = 14.2%, than trials in the control condition, mean = 20.4%, *SD* = 8.2%; $t(39) = 2.89$, $p < .01$.

Discussion

The results obtained in Experiment 1 provide evidence that a visual speech signal facilitates identification of subsequently presented auditory speech when the latter is presented in noise. This result complements and extends previous results in the literature indicating that speech perception is not limited to the auditory modality (e.g., Bernstein, 2005; Massaro, 1987, 1998; Sumby & Pollack, 1954), and that visual speech can prime perception of auditory speech produced by the same event (Kim et al., 2004). We will return to a discussion of these broad issues in the general discussion.

More specifically, Experiment 1 provided critical evidence suggesting that observers who are performing a spoken word recognition task are able to use speech information presented in the visual modality to inform their perception of the subsequently presented auditory target; thus, the repetition priming effect is not modality-specific. This finding replicates the earlier results reported by Kim et al. (2004), who found asynchronous cross-modality priming in a lexical decision task as opposed to the spoken word recognition task employed here (also see Dodd et al., 1989). Building on their results, we explored an additional factor which further suggests the use of visual information in the spoken word recognition task: the content of the responses that differed from the target words. In this analysis, we found additional evidence that the information present in the visual signal influences all responses in the spoken word recognition task; even the incorrect responses were more likely to be correct with respect to place and manner of articulation.

In the second experiment, we explored whether the priming effect observed in Experiment 1 requires that the two signals come from the same exemplar (i.e., same episodic experience). In our view, if we find that the cross-modality repetition priming effect persists even when the visual and auditory signals come from different sources (i.e., different talkers), this suggests that the priming effect occurs at a level of abstract (i.e., nonepisodic) representations which are activated during speech perception in addition to the processing of the specific event itself.

EXPERIMENT 2

The second experiment extended the findings of Experiment 1 by presenting participants with trials in which the visual signal and the auditory signal were produced by different talkers, and hence were from different speech events. The experimental manipulation allowed us to determine the level at which the repetition priming effect obtained in Experiment 1 occurred. In particular, we tested the possibility that this effect occurs at an instance-specific (or episodic) level of lexical representation. If this is the case, we would expect the priming effect to be attenuated when the visual speech video clip was produced by a different speech event than the spoken word target. In contrast, a lack of attenuation of the priming effect is expected if the priming arises at a level of lexical representation which is not instance-specific but rather that encodes phonetic structure that is abstracted over the exemplars one has encountered.

Method

Participants—Twenty-six Indiana University undergraduate students, ages 18–23, participated in Experiment 2. All participants were native speakers of English with no speech or hearing disorders. Participants received either course credit or monetary compensation for their participation in this study. None of the participants from Experiment 2 had participated in Experiment 1.

Materials—All stimulus materials were drawn from the Hoosier multi-talker audio-visual (AV) database (Sheffert et al., 1997). Monosyllabic, CVC words produced by the same female speaker and male speaker as in Experiment 1 were selected for this study. In Experiment 2, a set of 240 different word tokens were used (see Appendix for stimulus list). As in Experiment 1, half of the stimuli were ‘Easy’ words – high frequency words in sparse phonological neighbourhoods (e.g., ‘fool’), while the other half were ‘Hard’ words – low frequency lexical items in high density neighbourhoods (e.g., ‘hag’; Luce & Pisoni, 1998).

Procedure—The testing situation was identical to that used in Experiment 1. Each participant was presented with eight different trial types, with all permutations of prime type (Visual speech vs. Control), prime gender (Female vs. Male), and target gender (Female vs. Male). The experimental trials were analysed as two groups: AV Matched (Female prime and Female target; Male prime and Male target) and AV Mismatched (Female prime and Male target; Male prime and Female target).

Results

Word identification accuracy—Data from Experiment 2 were analysed with a 2 × 2 × 2 Prime type (Visual Speech/Control) vs. Target type (Easy/Hard) vs. AV-matching (Matched/Mismatched) ANOVA. Consistent with the results reported from Experiment 1, the results indicated a significant main effect of Prime type; words from Visual speech prime trials were identified more accurately, mean = 65.6%, *SD* = 10.8%, than words from Control prime trials, mean = 54.4%, *SD* = 11.6%; $F_1(1, 25) = 108.3$, $p < .001$; $F_2(1, 119) = 49.1$, $p < .001$. A significant main effect of Target type was also observed, with Easy targets recognised more accurately, mean = 65.1%, *SD* = 11.1%, than Hard Targets, mean = 54.9%, *SD* = 11.9%; $F_1(1, 25) = 85.6$, $p < .001$; $F_2(1, 119) = 23.8$, $p < .001$. No significant main effect was found for AV Matching, $F_1(1, 25) = 0.9$, *ns*; $F_2(1, 119) = 0.9$, *ns*, reflecting the lack of a difference in overall accuracy on AV-Matched and AV-Mismatched trials, regardless of Prime or Target type. Critical planned comparisons examined effects of Prime type separately for AV-Matched and AV-Mismatched trials. These comparisons revealed a significant effect of Prime type for both Matched: Visual speech, mean = 66.6%; Control, mean = 55.1%; $t(25) = 3.27$, $p < .01$; and Mismatched: Visual speech, mean = 64.4%;

Control, mean = 54.4%; $t(25) = 4.01$, $p < .001$, revealing that the spoken word recognition priming effect observed in the single-speaker condition does not crucially rely on the signals in the two stimulus presentation modalities coming from the same source. No significant interactions were obtained in the two-speaker conditions (all *F*s < 1.6).

Response analysis—We performed the same analyses on the set of responses in Experiment 2 as we did in Experiment 1. Collapsing over all the data, there were 3120 responses to targets with Visual speech primes and 3120 responses to targets with Control primes. A total of 1010 unique responses were given for Visual speech trials, whereas 1180 unique responses were given for Control trials. A chi-square analysis revealed that there were significantly more unique responses to Control trials than to Visual speech trials, $\chi^2(1) = 19.70$, $p < .01$. This finding strengthens the results reported in Experiment 1, indicating that the information present in the Visual speech prime acts as a constraint on the participants' responses to the auditory word presented in noise. When we examined the number of unique responses on the 1620 Matched Visual speech trials (659 unique responses) and the 1620 Mismatched Visual speech trials (700 unique responses), there was no significant difference between these two groups, $\chi^2(1) = 1.98$, ns, indicating that there was no difference in the constraint on responses for these conditions reflected by the number of unique responses for Matched trials and for Mismatched trials. Overall, more unique responses were produced on trials with Hard targets (893 unique responses) than on trials with Easy targets, 769 unique responses; $\chi^2(1) = 18.59$, $p < .01$. No significant differences were found in the proportion of unique responses to Easy and Hard words for any of the priming conditions (Visual speech Matched: Easy – 309, Hard – 350; Visual speech Mismatched: Easy – 324, Hard – 376; Control: Easy – 524, Hard – 638).

Following the analyses used in Experiment 1, each response was coded for the number of correct segments of the CVC word (i.e., 0–3 segments correct), and the average number of correct segments for each participant was computed for each condition. A repeated measures ANOVA revealed a significant main effect of Prime type, $F(1, 25) = 69.26$, $p < .001$, on the number of correct segments, with responses on trials with Visual speech primes (mean = 2.46, *SD* = 0.15) having significantly more correct segments than responses on trials with Control primes (mean = 2.27, *SD* = 0.14). The difference between the Matched (mean = 2.48, *SD* = 0.16) and the Mismatched (mean = 2.43, *SD* = 0.17) groups approached – but did not reach – significance, $t(25) = 2.03$, $p < .06$. When compared with the static trials, performance was significantly better for both Matched, $t(25) = 8.41$, $p < .001$, and Mismatched, $t(25) = 6.72$, $p < .001$, trials. These data provide further support for the claim that the responses are constrained by the presence of the linguistic information available in the Visual speech primes. This ANOVA also revealed a significant main effect of Target type, $F(1, 25) = 34.71$, $p < .001$, with responses on trials with Easy targets having more segments correct (mean = 2.41, *SD* = 0.17) than responses on trials with Hard targets (mean = 2.30, *SD* = 0.19). There was no significant interaction between Prime type and Target type.

When the analysis was limited to responses in which the participant gave the wrong whole word response, a repeated measures ANOVA revealed a significant main effect of Prime type, $F(1, 25) = 14.10$, $p < .05$, with the number of correct segments on trials with Visual speech primes (mean = 1.43, *SD* = 0.28) significantly greater than the number of correct segments on trials with Control primes (mean = 1.36, *SD* = 0.18). There was no significant difference between performance on Matched (mean = 1.44, *SD* = 0.21) and Mismatched (mean = 1.42, *SD* = 0.16) trials, $t(25) = 0.69$, *ns*. When compared with the number of segments correct in incorrect responses for the Control condition, there were significantly more segments correct for Visual speech trials in both the Matched, $t(25) = 2.34$, $p < .05$, and Mismatched, $t(25) = 2.10$, $p < .05$, AV conditions. This latter result confirms again that the phonetic information present in the Visual speech primes constrains all of the participants' responses leading to greater accuracy even for incorrect responses, and that this effect is not attenuated by having a gender mismatch between the source of the Visual speech prime and the auditory target.

The ANOVA also revealed a significant main effect of Target type, $F(1, 25) = 26.39$, $p < .05$, with the number of segments correct in incorrect responses higher for Hard targets (mean = 1.48, $SD = 0.14$) than for Easy targets (mean = 1.31, $SD = 0.19$). As in Experiment 1, this reflects a greater number of neighbours given as responses for Hard targets, 807/1406, 57.4%, than for Easy targets, 521/1089, 47.8%; $\chi^2 = 22.12$, $p < .05$. The interaction between Prime type and Target type was not significant for Experiment 2.

Following the analyses in Experiment 1, we analysed the likelihood of correct responses for each syllable position of the CVC words as a function of Prime type (collapsing over Target types). These data are presented in Table 3, with Matched and Mismatched conditions listed separately as well as combined. These data reveal that the overall accuracy is increased for words in the Visual speech condition compared with the Control condition for each of the syllable positions, suggesting that the Visual speech information helped constrain responses for all three segments of the CVC words.

To determine whether there was a difference in the accuracy benefit for any of the three positions, we computed a difference score for each syllable position. Overall planned comparisons indicated that the cross-modality priming effect was significantly greater for onset position than it was for either nucleus position, $t(25) = 3.07$, $p < .01$, or for coda position, $t(25) = 3.46$, $p < .01$, but there was no difference between accuracy on the nucleus position and coda position, $t(25) = 0.88$, *ns*. Comparisons limited to Matched and Mismatched Visual speech trials exhibit the same pattern, with onset position having significantly greater priming benefit than nucleus or coda, and with no significant difference observed between nucleus and coda.

The analyses of the data from Experiment 2 presented thus far suggest that there was a global benefit from the Visual speech information which constrained all components of the participants' responses. Further, these effects were observed even when there was neither temporal synchrony nor source identity of the auditory and Visual speech video speech signals. As discussed above, it is critical to investigate whether the priming benefit reflects a general benefit from the information present in the video clip, or whether the responses are constrained by the stimulus by limiting specific components of the set of competing hypotheses about the target word.

Following the analyses in Experiment 1, we examined the likelihood that participants would correctly identify particular phonological properties of the target stimulus. In particular, we examined the likelihood that participants would correctly identify the place features, manner features, and voicing features of the onset and coda consonants in the target word. The results are presented in Table 4.

The data in Table 4 indicate that the visual speech primes promote a robust increase in accuracy with respect to place and manner of articulation for onset and coda consonants. Crucially, these effects hold for both Matched and Mismatched primes; that is, the responses were significantly more accurate for both place and manner features even when the prime and target came from a different source. The performance on Matched and Mismatched trials did not differ significantly for any comparisons in Table 4 other than Onset place, where the identification of place for Matched trials was significantly better than identification of place for Mismatched trials ($\chi^2 = 8.68$, $p < .05$). With respect to voicing, following the analyses in Experiment 1, we limited our analysis to those trials in which the target and response were obstruents and thus the voice feature would have to be specified as part of the response. As with Experiment 1, there was no significant effect of prime type on accuracy of the voice feature, even when the analyses are limited further to just those trials with incorrect responses in which both the target and response have obstruents. To augment this analysis, we examined the likelihood that incorrect responses were part of the same lexical equivalence class as the target, as in Experiment 1. A paired t-test indicated that trials with Visual speech primes were more likely to have incorrect responses that are part of the same lexical equivalence class as the target, mean = 26.5%, *SD* = 5.1%, than trials in the control condition, mean = 20.7%, *SD* = 7.2%; *t*(25) = 3.52, *p* < .01.

GENERAL DISCUSSION

The experiments reported in this paper used a version of the short-term repetition priming paradigm to address issues related to visual and auditory speech perception. Participants were required to identify degraded spoken words presented in envelope-shaped noise that were preceded by visual speech primes or a control prime. In Experiment 1, the results indicated that participants were more accurate at identifying spoken words when the auditory stimulus was preceded by a visual speech stimulus of the same word compared with a control condition. Furthermore, detailed analyses of the participants' responses indicated that the visual speech prime constrained the responses to the auditory target even on trials where spoken word recognition was not successful, revealing that the visual speech prime affected lexical access by increasing the likelihood that particular lexical items would be selected for response. In Experiment 2, the same priming benefit was observed even when the auditory and visual signals came from different speakers, and hence different instances of the lexical category.

The present set of results raise several issues regarding the nature of speech perception and the level of processing at which these repetition priming effects take place. We have demonstrated that cross-modality repetition priming with visual primes and auditory targets is a robust phenomenon which occurs even when the commonality that exists between the visual speech prime and auditory speech target was only at the level of the lexical identity of the token being produced, and not identity of the token or specific ‘episode’ that is being perceived. Although it has been shown that observers may perceive asynchronous auditory and visual signals as having a single episodic source (van Wassenhove, Grant, & Poeppel, 2007), we do not believe that this was possible in the present study as there was a lag of 500 ms between the *offset* of the visual stimulus and the onset of the auditory stimulus. The result reported here is consistent with a view of speech perception in which auditory and visual information are both used in the cognitive process(es) involved in speech perception (Bernstein, 2005; Hamilton, Shenton, & Coslett, 2006; Kim et al., 2004; Massaro & Stork, 1998).[3] According to this type of account, language users store and maintain in memory abstract, internal representations of the phonetic components of speech, such as a representation of /p/. The results of the cross-modality repetition priming experiments reported here suggest that these representations may be activated directly by an acoustic waveform containing particular sounds (e.g., [p]), and (either directly or indirectly) by visual speech displays of a speaker creating the articulatory gestures that produce the same speech sounds (e.g., a labial closure).

[3] This type of theoretical approach posits that sensory information from the world is encoded in modality-specific representations, and that these modality-specific representations are either: (a) linked directly to one another (Massaro & Stork, 1998); or (b) linked to a separate ‘multimodal’ representation that integrates information from different sources (Hamilton et al., 2006; Skipper, Nusbaum, & Small, 2005). However, the difference between these proposals cannot be addressed by the research reported here.

The results reported here also reveal that the priming benefit observers received from the visual speech prime was under tight stimulus control. Five observations indicated that participants' spoken word recognition responses were highly constrained by the information present in the visual speech clip. First, more correct responses to auditory targets were observed on trials with visual speech primes. This result indicates that the *whole* target word was a more likely response on trials with visual speech primes. Second, across responses from all participants, we observed a smaller range of responses provided on trials with visual speech primes compared with the control primes.

Third, the presentation of the visual speech primes increased correct identification of the component segments in all three of the syllable positions of the CVC targets, with onsets benefiting more than the nucleus and coda. This pattern is consistent with a large body of literature suggesting that the initial segment or part of a word is psychologically important in a variety of perceptual and production tasks (Cole & Jakimik, 1980; Marslen-Wilson & Zwitserlood, 1989; Treiman, 1986; Treiman & Danis, 1988; Vitevitch, 2002), although word-final position has also consistently been found to create priming effects as well (e.g., Slowiaczek, Nusbaum, & Pisoni, 1987). It should be noted that a variety of factors could contribute to this result, including the smaller number of elements that can occupy nuclei and codas compared with onsets, the possibility that nuclei and codas are more closely coarticulated than onsets and nuclei leading the onsets to be more visually distinguishable, and variation in the amount of linguistic information conveyed in visual speech for each syllabic position due to the nature of the sounds being produced.

Fourth, the responses on trials with visual speech primes were more likely to exhibit accurate identification for two kinds of sub-segmental information: place of articulation and manner of articulation of both onset and coda consonants. This pattern suggests that there is reliable phonetic information about both place and manner of articulation in the visual speech video clip, providing another piece of evidence against the hypothesis that place information comes from the visual modality whereas manner information comes from the auditory modality (e.g., see Summerfield, 1987, for a statement and refutation of that proposal). In contrast, visual speech primes did not significantly increase the (already high) likelihood of accurately reporting the correct voicing status of the target obstruents. This may reflect that this component of the speech signal is not available in the visual speech stream and thus did not receive any benefit from the visual speech visual display, as has been argued elsewhere (e.g., Summerfield, 1979). However, it is worth noting that other cues to consonant voicing may be present in the visual speech signal, such as longer vowel durations before voiced consonants compared with voiceless consonants (Peterson & Lehiste, 1960). This result was augmented by the analyses indicating that the incorrect responses in the visual speech primes condition were more likely to share a lexical equivalence class (based on visual speech information) with the target than the responses in the control condition.

Finally, the results revealed an interaction between prime type (Visual speech vs. Control) and target type (‘Easy’ vs. ‘Hard’) with respect to the number of correct segments. This result indicates that both auditory and visual speech processing interact with processes used for lexical access, as responses to high-frequency words in sparse lexical neighbourhoods showed more benefit from the visual speech clip than responses to low-frequency words in dense lexical neighbourhoods. These data are consistent with earlier findings that visual and audiovisual speech processing interact with lexical access (Iverson, Bernstein, & Auer, 1998; Mattys et al., 2002) as well as the finding that auditory priming is stronger for words with fewer competitors (Dufour & Peereman, 2003).

In sum, the results presented above provide additional evidence suggesting that visual speech can prime the phonetic and lexical identification of auditory speech in a spoken word recognition task, and that this priming effect occurs at a level of representation that is further removed from the speech signals than modality-specific or episodic representations (see Sheffert & Fowler, 1995 for evidence that visual speech information may not be encoded in an exemplar lexicon).

One additional possibility that was not specifically addressed is that the participants may receive a benefit from the visual primes because they are covertly imitating the visual stimuli. This is consistent with the revised Motor theory of speech perception (Liberman & Mattingly, 1985) which suggests that encountering a speech signal leads to the recovery of the intended articulatory gestures, and there has been some evidence suggesting that this is an unavoidable consequence in visual speech perception (Kerzel & Bekkering, 2000). It is also possible that, given the 500 ms ISI, the participants are actively attempting to understand the prime word from the dynamic visual speech clip.[4] In addition, it is worth noting that it remains possible that the effects in the identification paradigm arise due to post-perceptual biases rather than as priming within the perceptual processes themselves. We believe that this possibility exists for all experimental paradigms that have been used to address these issues (e.g., lexical decision), and that the five lines of evidence suggesting that the responses are under tight stimulus control support the notion that the perceptual processes themselves are the locus of the effects reported here. In the remainder of this paper, we discuss two issues that are raised by the present study: indexical properties in audiovisual speech perception; and general properties of lexical access and lexical competition.

Cross-modal identity matching and talker-familiarity effects

The topic of multimodal speech perception has received attention from speech perception theorists (e.g., Fowler, 1996, 2004; Massaro, 1998) as well as researchers addressing a wide variety of problems including second language acquisition (Davis & Kim, 2001, 2004; Kim & Davis, 2003), neurological processes and impairment (Hamilton et al., 2006; Skipper et al., 2005), perception among hearing-impaired individuals (Bergeson & Pisoni, 2004; Lachs, Pisoni, & Kirk, 2001; Massaro & Cohen, 1999), speech production (Yehia, Rubin, & Vatikiotis-Bateson, 1998), voice identity (Kamachi, Hill, Lander, & Vatikiotis-Bateson, 2003; Lachs, 2002; Lachs & Pisoni, 2004a, 2004b), and issues directly related to spoken word recognition (Dodd et al., 1989; Kim et al., 2004; Mattys et al., 2002). These audiovisual speech studies – combined with the pioneering work of Sumby and Pollack (1954) – all reveal that a fundamental component of speech perception is the simultaneous use of information from auditory and visual signals.

One recent series of studies in the multimodal speech perception literature has revealed that perceivers are able to match a video of a speaker's face to the appropriate corresponding voice when visual and auditory stimuli are presented separately in time (Lachs, 2002; Lachs & Pisoni, 2004a, 2004b), similar to the presentation in the present study. This cross-modality matching task can be performed successfully even when the linguistic content of the two signals differs (Kamachi et al., 2003), suggesting that the perceptual cues used for

[4] Several subjects were run on the same experiment with an ISI of 50 ms, and there were no differences between their data and those reported above. Given the relatively short ISI in these other participants, it is unlikely that these subjects were able to rehearse and decode the target word from the visual speech prime. However, it is worth noting that these similar patterns could arise from different loci, with the participants in the experiments reported here rehearsing and decoding the target words, and the participants who saw the stimuli with a 50 ms ISI exhibiting facilitatory effects of auditory processing from the visual speech signals (as in van Wassenhove, Grant, & Poeppel, 2005). We thank an anonymous reviewer for articulating this issue.

cross-modality identity matching are independent of the idiosyncrasies of a particular utterance.

Lachs and Pisoni (2004a, 2004b) suggested that their participants' success in cross-modality identity matching – in which the correctly matched stimuli came from the same utterance – may be rooted in event-based perception (Gibson, 1966). Lachs and Pisoni's auditory and visual stimuli provided information about the same physical event in the world, and they argued that ‘integration’ of the two modalities of information came from the real-world event itself, which shaped and constrained the specific pattern of sensory stimulation impinging on the eyes and ears. Within the direct realist event-based theoretical framework (Fowler, 1986), acoustic and optical speech signals are integrated seamlessly because they are two sources specifying information about the same distal event (also see Fowler, 2004).

A similar conclusion was drawn recently by Rosenblum et al. (2007) who presented evidence indicating that subjects performed better on an auditory sentence transcription task after they were familiarised with a speaker in a visual-only condition. Rosenblum et al. argued that this effect arises because observers became familiar with amodal talker-specific properties from the visual speech block which enabled them to perform better in the auditory speech block. Similar effects of talker-familiarity have been reported in auditory speech perception (e.g., Nygaard, Sommers, & Pisoni, 1994; Palmeri, Goldinger, & Pisoni, 1993) as well as in visual and audiovisual speech perception (see Rosenblum, 2005 for a review).

The apparent difference between the results discussed by Rosenblum et al. (2007) and those reported here could be due to a difference in task (sentence transcription vs. single word recognition), amount of exposure to visual speech (Rosenblum et al. exposed participants to visual speech for approximately 50 minutes), or level of visual speech perception expertise among the participants (Rosenblum et al.'s participants correctly identified at least 32.5% of words in a visual speech screening task). It is also worth noting here that the McGurk effect (McGurk & MacDonald, 1976) is maintained even when there is a gender mismatch between the face producing the visual speech signal and the voice producing the auditory speech signal (Green, Kuhl, Meltzoff, & Stevens, 1991; also see Vatakis & Spence, 2007 for a detailed discussion of these issues), although there is a substantial difference between simultaneously presented stimuli in the McGurk illusion (or slightly asynchronous, as in Vakatis & Spence, 2007) and the sequentially presented visual and auditory signals in the present experiment

The results of Experiment 2 – which provided clear and consistent evidence indicating that the facilitatory effects of cross-modality priming are maintained even in a condition where there was a mismatch between the speakers – are consistent with an event-based perception account in which the event is defined with respect to the articulatory gestures that create the visual and auditory percept (e.g., [p] defined as voiceless labial stop), regardless of who produces them or whether they come from the same articulatory source. Given this definition of speech event, there is no reason to predict that the cross-modality priming effect would be absent when there is a lack of identity in the source of the two stimuli, and the apparent conflict between these results and the cross-modality identity matching results (e.g., Lachs & Pisoni, 2004a, 2004b) arises due to the nature of the tasks that are being performed. It is also well worth noting that in the present experiment, the visual and auditory signals were presented sequentially and it remains possible that these two sensory signals were not interpreted as a single event by the participants in the study.

Open-set identification and lexical access

One additional finding which emerged from this study provides further insight into the nature of lexical competition in the process of lexical access regardless of input modality.

For both experiments reported here, when we analysed the incorrect whole word responses (i.e., failures of lexical access), we observed more correct segments on trials with ‘hard’ target words (i.e., low frequency words from dense lexical neighbourhoods) than on trials with ‘easy’ target words (high frequency words from sparse lexical neighbour-hoods). This finding was largely attributable to a larger number of incorrect responses with two segments correct on trials with hard targets than on trials with easy targets. The definition of lexical neighbour used in this paper, based on Luce and Pisoni (1998), was a word that shares all but one segment with the target word. Thus, it was more likely that incorrect responses for ‘hard’ targets were neighbours of the target word (i.e., sharing two of the three segments) than that incorrect responses for ‘easy’ targets were neighbours of the target word. While this result follows from the Neighbourhood Activation Model (NAM) of Luce and Pisoni (1998) in a straightforward manner, it is a novel empirical demonstration of a critical component of NAM.

The Neighbourhood Activation Model assumes that the strength and number of competitors directly influences the ease with which lexical items are recognised in any task that requires lexical access (Luce & Pisoni, 1998). Previous attempts to understand the role of neighbourhood density in spoken word recognition have typically focused on measures of accuracy and processing time, and it has been repeatedly observed (Luce & Pisoni, 1998; Vitevitch & Luce, 1998, 1999; Vitevitch, Luce, Pisoni, & Auer, 1999) that words with strong competitors (i.e., ‘hard’ words) are processed more slowly and less accurately than words with weaker competitors (i.e., ‘easy’ words). However, previous reports have not included detailed analyses of the error responses as were presented in this paper. The results reported here provide further support for the fundamental claim underlying NAM by demonstrating that when lexical access fails, the response is more likely to be a lexically similar neighbour/competitor for ‘hard’ words than it is for ‘easy’ words.

Conclusion

The results from two cross-modality repetition priming experiments indicated that identification of spoken words mixed with noise was facilitated by the earlier presentation of a visual speech clip of the same lexical item. The present set of findings indicates that visual speech perception and auditory speech perception rely on (at least) some shared cognitive resources, and that the repetition priming effects occur at a level of representation that is neither modality-specific nor instance-specific. The cross-modality repetition priming paradigm can be used in the future to provide critical new information pertaining to the nature of speech perception by exploring the nature of the stimuli that produce this effect. We expect that these lines of research will help elucidate the neural mechanisms underlying repetition priming and repetition suppression (see Grill-Spector, Henson, & Martin, 2006 for a recent review) in speech perception and be extended to multimodal perception (e.g., see Ghazanfar & Schroeder, 2006). In addition, these lines of research are directly relevant to understanding the relation of the two input modalities in clinical populations such as hearing-impaired listeners who have experienced a period of auditory deprivation that may encourage reorganisation and remodelling of the typical developmental processes (Bergeson & Pisoni, 2004; Lachs et al., 2001).

Acknowledgments

This work was supported by NIH DC00012. The authors would like to thank Melissa Troyer for her assistance with this study, and Susannah Levi, Tessa Bent, Manuel Carreiras, and several anonymous reviewers for helpful comments on earlier drafts of this paper.

APPENDIX

Experimental word stimuli. Stimuli listed in **bold** used in both Experiment 1 and Experiment 2. All other stimuli were used exclusively in Experiment 2.

back
badge
bait
bake
ban
bang
base
beach
bean
bed
beer
boat
bone
boot
bore
both
bud
bug
bum
cake
call
cane
case
cat
cause
cave
chain
chair
chat
check
cheer
cheese
chief
chin
chore
church
cite
coat
cod
comb
con
cone
cool
curve
dam
dame
dare
date
deal
death
debt
deep
den
dig
dirt
does
dog
doom
doubt
down
dune
face
fade
fair
faith
fall
fan
fat
fear
feel
fig
fine
fire
firm
fit
five
fool
full
gain
gas
gave
girl
give
goat
god
gown
guide
gum
gut
hag
ham
hash
hen
hick
hike
hood
hoot
hope
house
hung
hurl
jack
job
join
judge
keep
king
kiss
knead
knob
knot
known
lace
lad
lake
lame
late
lawn
learn
leave
leg
less
lice
light
love
luck
mail
main
map
mat
meat
mile
mine
mitt
moat
mole
mood
mouse
mouth
mum
neck
net
noise
pace
pad
page
pail
pain
pan
pat
path
pawn
peace
pen
pet
pick
pin
pool
pope
pot
pup
push
rang
rat
rate
reach
real
reed
rich
ring
rise
road
roar
rock
roof
root
rose
rough
rule
rum
rut
scene
seat
seek
serve
shade
shape
shed
sheet
shell
shop
shore
sick
sign
sill
size
soil
south
suck
tack

loan	nose	rain	sad	take
long	note	raise	sail	talk
loose	one	**rake**	sane	taught
teat	**tin**	**vice**	**wash**	**work**
teeth	**ton**	**voice**	**watch**	
theme	**toot**	**wad**	weak	
thick	top	**wade**	**wed**	
thumb	town	wail	white	
tile	use	was	**wife**	

REFERENCES

Auer ET Jr. The influence of the lexicon on speech read word recognition: Contrasting segmental and lexical distinctiveness. Psychonomic Bulletin and Review. 2002; 9(2):341–347. [PubMed: 12120798]

Auer ET Jr. Bernstein LE. Speechreading and the structure of the lexicon: Computationally modeling the effects of reduced phonetic distinctiveness on lexical uniqueness. Journal of the Acoustical Society of America. 1997; 102(6):3704–3710. [PubMed: 9407662]

Bergeson, TR.; Pisoni, DB. Audiovisual speech perception in deaf adults and children following cochlear implantation. In: Calvert, GA.; Spence, C.; Stein, BE., editors. The handbook of multisensory processes. MIT Press; Cambridge, MA: 2004.

Bernstein, LE. Phonetic processing by the speech perceiving brain. In: Pisoni, DB.; Remez, RE., editors. Handbook of speech perception. Blackwell; Malden, MA: 2005. p. 79-98.

Bowers, JS.; Marsolek, CJ., editors. Rethinking implicit memory. Oxford University Press; New York: 2003.

Bybee, J. Frequency and language use. Cambridge University Press; Cambridge: 2001.

Calvert, GA.; Spence, C.; Stein, BE., editors. The handbook of multisensory processes. MIT Press; Cambridge, MA: 2004.

Cole, RA.; Jakimik, J. A model of speech perception. In: Cole, RA., editor. Perception and production of fluent speech. Lawrence Erlbaum Associates; Hillsdale, NJ: 1980. p. 133-163.

Davis C, Kim J. Repeating and remembering foreign language words: Implications for language teaching systems. Artificial Intelligence Review. 2001; 16:37–47.

Davis C, Kim J. Audio-visual interactions with intact clearly audible speech. Quarterly Journal of Experimental Psychology. 2004; 57A(6):1103–1121. [PubMed: 15370518]

Dodd B, Oerlemens M, Robinson R. Cross-modal effects in repetition priming: A comparison of lip-read graphic and heard stimuli. Visible Language. 1989; 22:59–77.

Dufour S, Peereman R. Inhibitory priming effects in auditory word recognition: When the target's competitors conflict with the word prime. Cognition. 2003; 88:B33–B44. [PubMed: 12804820]

Foster KI, Davis C. Repetition priming and frequency attenuation in lexical access. Journal of Experimental Psychology: Learning, Memory and Cognition. 1984; 10:680–689.

Fowler C. An event approach to the study of speech perception from a direct-realist perspective. Journal of Phonetics. 1986; 14:3–28.

Fowler C. Listeners do hear sounds, not tongues. Journal of the Acoustical Society of America. 1996; 99:1730–1741. [PubMed: 8819862]

Fowler, C. Speech as a supramodal or amodal phenomenon. In: Calvert, GA.; Spence, C.; Stein, BE., editors. The handbook of multisensory processes. MIT Press; Cambridge, MA: 2004.

Ghazanfar AA, Schroeder CE. Is neocortex essentially multisensory? Trends in Cognitive Science. 2006; 10:278–285.

Gibson, JJ. The senses considered as perceptual systems. Houghton Mifflin; Boston, MA: 1966.

Goldinger SD. Words and voices: episodic traces in spoken word identification and recognition memory. Journal of Experimental Psychology: Learning, Memory and Cognition. 1996; 22:1166–1183.

Goldinger SD. Echoes of echoes? An episodic theory of lexical access. Psychological Review. 1998; 105(2):251–279. [PubMed: 9577239]

Goldinger SD, Azuma T. Puzzle-solving science: The quixotic quest for units in speech perception. Journal of Phonetics. 2003; 31:305–320.

Green KP, Kuhl PK, Meltzoff AN, Stevens EB. Integrating speech information across talkers, gender, and sensory modality: Female faces and male voices in the McGurk effect. Perception and Psychophysics. 1991; 38:269–276. [PubMed: 4088819]

Grill-Spector K, Henson R, Martin A. Repetition and the brain: Neural models of stimulus-specific effects. Trends in Cognitive Science. 2006; 10(1):14–23.

Hamilton RH, Shenton JT, Coslett HB. An acquired deficit of audiovisual speech processing. Brain and Language. 2006; 98:66–73. [PubMed: 16600357]

Horii Y, House AS, Hughes GW. A masking noise with speech envelope characteristics for studying intelligibility. Journal of the Acoustical Society of America. 1971; 49:1849–1856. [PubMed: 5125732]

Huttenlocher, DP.; Zue, VW. A model of lexical access from partial phonetic information; Paper presented at the IEEE International Conference on Acoustics, Speech and Signal Processing; 1984.

Iverson P, Bernstein LE, Auer ET Jr. Modeling the interaction of phonemic intelligibility and lexical structure in audiovisual word recognition. Speech Communication. 1998; 26:45–63.

Johnson, K. Speech perception without speaker normalization: an exemplar model. In: Johnson, K.; Mullenix, JW., editors. Talker variability in speech processing. Academic Press; San Diego, CA: 1997. p. 145-166.

Johnson, K. Decisions and mechanisms in exemplar-based phonology. UC Berkeley; Berkeley, CA: 2005.

Kamachi M, Hill H, Lander K, Vatikiotis-Bateson E. ‘Putting the face to the voice’: Matching identity across modality. Current Biology. 2003; 13:1709–1714. [PubMed: 14521837]

Kerzel D, Bekkering H. Motor activation from visible speech: Evidence from stimulus response compatibility. Journal of Experimental Psychology: Human Perception and Performance. 2000; 26(2):634–647. [PubMed: 10811167]

Kim J, Davis C. Task effects in masked cross-script translation and phonological priming. Journal of Memory and Language. 2003; 49:484–499.

Kim J, Davis C, Krins P. Amodal processing of visual speech as revealed by priming. Cognition. 2004; 93(1):B39–B47. [PubMed: 15110729]

Lachs, L., editor. Vocal tract kinematics and crossmodal speech information. Speech Research Laboratory, Indiana University; Bloomington, IN: 2002.

Lachs, L.; Hernandez, LR. Research on spoken language processing progress report no. 22. Speech Research Laboratory, Indiana University; Bloomington, IN: 1998. Update: The Hoosier audiovisual multi-talker database; p. 377-388.

Lachs L, Pisoni DB. Cross-modal source information and spoken word recognition. Journal of Experimental Psychology: Human Perception and Performance. 2004a; 30(2):378–396. [PubMed: 15053696]

Lachs L, Pisoni DB. Crossmodal source identification in speech perception. Ecological Psychology. 2004b; 16(3):159–187.

Lachs L, Pisoni DB, Kirk KI. Use of audiovisual information in speech perception by prelingually deaf children with cochlear implants: A first report. Ear and Hearing. 2001; 22:236–251. [PubMed: 11409859]

Lachs L, Weiss JW, Pisoni DB. Use of partial stimulus information by cochlear implant users and listeners with normal hearing in identifying spoken words: Some preliminary analyses. The Volta Review. 2000; 102(4):303–320.

Liberman AM, Mattingly IG. The motor theory of speech perception revised. Cognition. 1985; 21:1–36. [PubMed: 4075760]

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

Marslen-Wilson WD, Zwitserlood P. Accessing spoken words: The importance of word onsets. Journal of Experimental Psychology: Human Perception and Performance. 1989; 15:576–585.

Massaro, DW. Speech perception by ear and eye. In: Dodd, B.; Campbell, R., editors. Hearing by eye: The psychology of lip-reading. Lawrence Erlbaum Associates; Hillsdale, NJ: 1987. p. 53-84.

Massaro, DW. Perceiving talking faces: From speech perception to a behavioral principle. MIT Press; Cambridge, MA: 1998.

Massaro DW, Cohen MM. Perceiving talking faces. Current Directions in Psychological Science. 1995; 4:104–109.

Massaro DW, Cohen MM. Speech perception in hearing-impaired perceivers: Synergy of multiple modalities. Journal of Speech, Language and Hearing Research. 1999; 42:21–41.

Massaro DW, Stork DG. Speech recognition and sensory integration: a 240-year-old theorem helps explain how people and machines can integrate auditory and visual information to understand speech. American Scientist. 1998; 86:236–244.

Mattys SL, Bernstein LE, Auer ET Jr. Stimulus-based lexical distinctiveness as a general word-recognition mechanism. Perception and Psychophysics. 2002; 64(4):667–679. [PubMed: 12132766]

McGurk H, MacDonald J. Hearing lips and seeing voices. Nature. 1976; 264(5588):746–748. [PubMed: 1012311]

McLennan CT, Luce PA, Charles-Luce J. Representation of lexical form. Journal of Experimental Psychology: Learning, Memory and Cognition. 2003; 29(4):539–553.

Miller GA, Nicely P. An analysis of perceptual confusions among some English consonants. Journal of the Acoustical Society of America. 1955; 27(2):338–352.

Morton, J. Word recognition. In: Morton, J.; Marshall, JC., editors. Structures and processes. MIT Press; Cambridge: 1979. p. 109-156.

Nygaard LC, Sommers MS, Pisoni DB. Speech perception as a talker-contingent process. Psychological Science. 1994; 5(1):42–46. [PubMed: 21526138]

Oldfield RC. Things, words, and the brain. Quarterly Journal of Experimental Psychology. 1966; 18:340–353. [PubMed: 5956077]

Palmeri TJ, Goldinger SD, Pisoni DB. Episodic encoding of voice attributes and recognition memory for spoken words. Journal of Experimental Psychology: Learning, Memory and Cognition. 1993; 19:309–328.

Peterson GE, Lehiste I. Duration of syllable nuclei in English. Journal of the Acoustical Society of America. 1960; 32(6):693–703.

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

Pisoni, DB.; Levi, SV. Representations and representational specificity in speech perception and spoken word recognition. In: Gaskell, MG., editor. Handbook of psycholinguistics. Oxford University Press; Oxford: in press

Port R, Leary A. Against formal phonology. Language. 2005; 81(4):927–964.

Ronquest, RE.; Levi, SV.; Pisoni, DB. Research on Spoken Language Processing Report No. 28. Speech Research Laboratory, Indiana University; Bloomington, IN: 2007. Language identification from visual-only speech; p. 95-118.

Rosenblum, LD. Primacy of multimodal speech perception. In: Pisoni, DB.; Remez, RE., editors. Handbook of speech perception. Blackwell; Malden, MA: 2005. p. 51-78.

Rosenblum LD, Miller RM, Sanchez K. Lip-read me now, hear me better later: Cross-modal transfer of talker-familiarity effects. Psychological Science. 2007; 18(5):392–396. [PubMed: 17576277]

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

Sheffert, SM.; Lachs, L.; Hernandez, LR. Research on spoken language processing progress report no. 21. Speech Research Laboratory, Indiana University; Bloomington, IN: 1997. The Hoosier audiovisual multi-talker database; p. 578-583.

Shipman, DW.; Zue, VW. Properties of large lexicons: Implications for advanced isolated word recognition systems; Paper presented at the IEEE 1982 International Conference on Acoustics, Speech and Signal Processing; 1982.

Skipper JI, Nusbaum HC, Small SL. Listening to talking faces: Motor cortical activation during speech perception. Neuroimage. 2005; 25:76–89. [PubMed: 15734345]

Slowiaczek LM, Nusbaum HC, Pisoni DB. Phonological priming in auditory word recognition. Journal of Experimental Psychology: Learning, Memory and Cognition. 1987; 13(1):64–75.

Soto-Faraco S, Navarra J, Weikum WM, Vouloumanos A, Sebastián-Gallés N, Werker JF. Discriminating languages by speech reading. Perception and Psychophysics. 2007; 69(2):218–237. [PubMed: 17557592]

Sumby WH, Pollack I. Visual contribution to speech intelligibility in noise. Journal of the Acoustical Society of America. 1954; 26:212–215.

Summerfield AQ. Use of visual information in phonetic perception. Phonetica. 1979; 36:314–331. [PubMed: 523520]

Summerfield, AQ. Some preliminaries to a comprehensive account of audio-visual speech perception. In: Dodd, B.; Campbell, R., editors. Hearing by eye: The psychology of lip-reading. Lawrence Erlbaum Associates; Hillsdale, NJ: 1987. p. 3-52.

Treiman R. The division between onsets and rimes in English syllables. Journal of Memory and Language. 1986; 25:476–491.

Treiman R, Danis C. Short-term memory errors for spoken syllables are affected by the linguistic structure of the syllables. Journal of Experimental Psychology: Learning, Memory and Cognition. 1988; 14:145–152.

van Wassenhove V, Grant KW, Poeppel D. Visual speech speeds up the neural processing of auditory speech. Proceedings of the National Academy of Sciences. 2005; 102(4):1181–1186.

van Wassenhove V, Grant KW, Poeppel D. Temporal window of integration in auditory-visual speech perception. Neuropsychologia. 2007; 45:598–607. [PubMed: 16530232]

Vatakis A, Spence C. Crossmodal binding: Evaluating the ‘unity assumption’ using audiovisual speech stimuli. Perception & Psychophysics. 2007; 69:744–756. [PubMed: 17929697]

Vitevitch MS. Influence of onset density on spoken word recognition. Journal of Experimental Psychology: Human Perception and Performance. 2002; 28(2):270–278. [PubMed: 11999854]

Vitevitch MS, Luce PA. When words compete: Levels of processing in spoken word perception. Psychological Science. 1998; 9:325–329.

Vitevitch MS, Luce PA. Probabilistic phonotactics and neighborhood activation in spoken word recognition. Journal of Memory and Language. 1999; 40:374–408.

Vitevitch MS, Luce PA, Pisoni DB, Auer ET Jr. Phonotactics, neighborhood activation and lexical access for spoken words. Brain and Language. 1999; 68:306–311. [PubMed: 10433774]

Weikum WM, Vouloumanos A, Navarra J, Soto-Faraco S, Sebastián-Gallés N, Werker JF. Visual language discrimination in infancy. Science. 2007; 316(5828):1159. [PubMed: 17525331]

Yehia H, Rubin P, Vatikiotis-Bateson E. Quantitative association of vocal-tract and facial behavior. Speech Communication. 1998; 26:23–43.

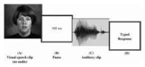

Figure 1.
Schematic of experimental trial. In Experiment 1, the video clip (A) and auditory clip (C) come from the same token of a single speaker. In Experiment 2, (A) and (C) come from the same speaker or from different speakers producing the identical lexical item.

TABLE 1

Response accuracy for each of the three syllable positions in the CVC stimuli as a function of prime type (Experiment 1).

	Visual speech % SD	Control % SD	Analysis
Onset	80.3 (8.3)	67.6 (7.4)	*t*(39)=9.18, *p* <.001
Nucleus	87.1 (6.1)	78.4 (7.3)	*t*(39)=6.38, *p* <.001
Coda	81.3 (7.9)	74.8 (8.3)	*t*(39)=4.39, *p* <.001

TABLE 2

Response accuracy assessed for sub-phonemic features for onset and coda consonants as a function of prime type (Experiment 1). The voicing data only examine obstruents as these are the only segments for which an error can be made.

		Visual speech % correct	Control % correct	Analysis
Place	Onset	86	76	$\chi^2(1)=52.66$, *p* <.001
	Coda	90	85	$\chi^2(1)=19.45$, *p* <.001
Manner	Onset	88	79	$\chi^2(1)=58.89$, *p* <.001
	Coda	89	86	$\chi^2(1)=7.28$, *p* <.01
Voice	Onset	98	97	$\chi^2(1)=2.28$, *ns*
	Coda	97	95	$\chi^2(1)=2.25$, *ns*

TABLE 3

Response accuracy for each of the three syllable positions in the CVC stimuli as a function of prime type (Experiment 2). Matched and Mismatched Visual speech trials are compared with overall data from Control trials.

		Visual speech % SD	Control % SD	Analysis
Onset	Total	79.9 (6.7)	(5.0)	*t*(25)=8.35, *p* <.001
	Matched	81.3 (6.3)		*t*(25)=9.76, *p* <.001
	Mismatched	78.3 (8.4)		*t*(25)=5.40, *p* <.001
Nucleus	Total	84.6 (6.1)	78.6 (5.5)	*t*(25)=5.82, *p* <.001
	Matched	85.4 (6.4)		*t*(25)=6.01, *p* <.001
	Mismatched	83.7 (6.0)		*t*(25)=4.25, *p* <.001
Coda	Total	81.7 (4.9)	76.8 (4.6)	*t*(25)=4.86, *p* <.001
	Matched	81.8 (6.1)		*t*(25)=4.19, *p* <.001
	Mismatched	81.5 (5.3)		*t*(25)=4.02, *p* <.001

TABLE 4

Accuracy in identifying the place, manner, and voice for onset and coda consonants. Statistical analyses compare the performance on static trials to performance on total Visual speech trials, as well as to Matched and Mismatched trials separately. The voicing data only examine obstruents as these are the only segments for which an error can be made.

Feature	Position	Prime type	Visual speech %	Control %	Analysis
Place	Onset	Total	86	78	$\chi^2(1)=52.3$, p <.001
		Matched	87		$\chi^2(1)=54.6$, p <.001
		Mismatched	84		$\chi^2(1)=17.2$, p <.001
	Coda	Total	89	85	$\chi^2(1)=26.0$, p <.001
		Matched	89		$\chi^2(1)=13.8$, p <.001
		Mismatched	89		$\chi^2(1)=19.4$, p <.001
Manner	Onset	Total	88	83	$\chi^2(1)=29.7$, p <.01
		Matched	89		$\chi^2(1)=26.7$, p <.01
		Mismatched	87		$\chi^2(1)=12.4$, p <.01
	Coda	Total	90	88	$\chi^2(1)=8.18$, p <.05
		Matched	90		$\chi^2(1)=6.48$, p <.05
		Mismatched	90		$\chi^2(1)=4.10$, p <.05
Voice	Onset	Total	98	97	$\chi^2(1)=3.08$, *ns*
		Matched	98		$\chi^2(1)=0.72$, *ns*
		Mismatched	98		$\chi^2(1)=2.47$, *ns*
	Coda	Total	95	94	$\chi^2(1)=1.80$, *ns*
		Matched	95		$\chi^2(1)=3.55$, *ns*
		Mismatched	95		$\chi^2(1)=1.96$, *ns*