Attention, Perception, & Psychophysics
https://doi.org/10.3758/s13414-023-02689-2

Semantic priming from McGurk words: Priming depends on perception

Josh Dorsi[1,2] · **Rachel Ostrand**[3] · **Lawrence D. Rosenblum**[1]

Accepted: 23 February 2023

Abstract
The McGurk effect is an illusion in which visible articulations alter the perception of auditory speech (e.g., video 'da' dubbed with audio 'ba' may be *heard* as 'da'). To test the timing of the multisensory processes that underlie the McGurk effect, Ostrand et al. *Cognition 151*, 96–107, 2016 used incongruent stimuli, such as auditory 'bait' + visual 'date' as primes in a lexical decision task. These authors reported that the auditory word, but not the perceived (visual) word, induced semantic priming, suggesting that the auditory signal alone can provide the input for lexical access, before multisensory integration is complete. Here, we conceptually replicate the design of Ostrand et al. (2016), using different stimuli chosen to optimize the success of the McGurk illusion. In contrast to the results of Ostrand et al. (2016), we find that the perceived (i.e., visual) word of the incongruent stimulus usually induced semantic priming. We further find that the strength of this priming corresponded to the magnitude of the McGurk effect for each word combination. These findings suggest, in contrast to the findings of Ostrand et al. (2016), that lexical access makes use of integrated multisensory information which is perceived by the listener. These findings further suggest that which unimodal signal of a multisensory stimulus is used in lexical access is dependent on the perception of that stimulus.

Keywords Multisensory processing · Semantic priming

Speech perception is inherently multisensory. Seeing the articulations of a talker can enhance perception of auditory speech, whether degraded by noise or foreign accent, or even if the speech is clear, but has technical content (e.g., Arnold & Hill, 2001; Reisberg et al., 1987; Sumby & Pollack, 1954). Regardless of one's level of hearing, *visual speech perception* is also used during first and second language acquisition (Navarra & Soto-Faraco, 2007; Teinonen et al., 2008), and influences the degree of phonetic alignment between interlocutors (e.g., Dias & Rosenblum, 2011). The multisensory nature of speech is also evidenced by neurophysiological research showing that the brain responds to auditory and visual speech input in remarkably similar ways (for a review, see Rosenblum et al., 2016a, b).

The most studied example of multisensory speech perception is the McGurk effect (McGurk & MacDonald, 1976; for reviews, see Alsius et al., 2018; Rosenblum, 2019). The McGurk effect is the finding that if visual speech segments are dubbed onto incongruent auditory segments, the results can be an illusory "heard" percept that differs from the auditory stimulus. For example, McGurk and MacDonald (1976) report that when auditory 'ba' is dubbed onto a visual 'ga,' perceivers report *hearing* either 'ga' (a visually dominated perception) or 'da' (a fusion perception). Since its discovery, the McGurk effect has been taken as a hallmark example of audiovisual integration (e.g., Bebko et al., 2014; Samuel & Lieblich, 2014; Stropahl et al., 2016; but see Alsius et al., 2018; Rosenblum, 2019).

The McGurk effect has also motivated much research on how multisensory integration fits into the overall language function. This research has provided both neurophysiological and behavioral data. Much of the neurophysiological work indicates that audiovisual integration occurs early in speech processing (for a review, see Rosenblum, Dias, et al., 2016). For example, audio-visual speech has been found to modulate auditory-evoked brainstem responses as early as 11 ms following acoustic stimulus onset (Musacchia et al.,

✉ Josh Dorsi
jdorsi@pennstatehealth.psu.edu

[1] Department of Psychology, University of California, Riverside, 900 University Ave, Riverside, CA 92521, USA

[2] Penn State University, College of Medicine, State College, PA, USA

[3] IBM Research, Yorktown Heights, NY, USA

2006). This result is consistent with the finding that visual speech produces activity in the auditory cortex (Calvert et al., 1997; Pekkola et al., 2005) as early as 10 ms following activation of the visual cortex (Besle et al., 2008). Finally, while likely the result of feedback interactions, visual speech can influence auditory processing in the cochlea as demonstrated by influences on transient-evoked otoacoustic emissions (Namasivayam et al., 2015). Collectively, these neurophysiological findings support the contention that audiovisual integration begins at the earliest stages of speech processing.

There are also behavioral studies which suggest that multisensory integration occurs very early in linguistic processing. For example, Green and Miller (1985) found that visual speech could produce a McGurk effect that influenced the perception of voice-onset-time (VOT; see also Brancazio & Miller, 2005; Green & Kuhl, 1989; Sanchez et al., 2010). As VOT is a prephonemic feature of speech perception, this finding suggests that audiovisual integration occurs prior to word, or even word segment, recovery. Similarly, the auditory perception of place of articulation for coarticulated speech is also sensitive to visual speech information (Fowler et al., 2000; Green & Norrix, 2001). These findings suggest that multisensory integration begins early, likely before segment features are established, and long before words are perceptually identified.

Many of these findings come from work with nonword syllable stimuli (such as ‘ba’ and ‘da’), which necessarily do not require lexical access. Therefore, it is possible that multisensory integration occurs later in the processing timeline for real-word stimuli, for which lexical processing *does* need to occur. Indeed, there is some evidence that lexical information influences the McGurk effect, suggesting that lexical access may occur *before* completion of multisensory integration. This work shows that the McGurk effect occurs more often when the illusory percept forms a real word as opposed to a nonword, suggesting that lexical knowledge influences audiovisual integration (Brancazio, 2004; but see Sams et al., 1998). Additionally, the degree of this McGurk integration is stronger when the audio-visually discrepant segment occurs in the word-final, as opposed to word-initial, position, suggesting that top-down lexical knowledge from the unfolding lexical input affects integration (e.g., Barutchu et al., 2008). Relatedly, the McGurk effect is also stronger when integration of the incongruent unimodal signals forms a word that is semantically consistent with the preceding sentence context, as compared with one which is not consistent (e.g.,Windmann, 2004). These findings suggest that there may be some interactivity between lexical access and multisensory integration.

Ostrand et al. (2016); see also Ostrand et al. (2011) investigated the relative timing of lexical access and multisensory integration using a semantic priming paradigm. On each trial, participants received an audiovisual prime followed by an audio-only target utterance, and performed a lexical decision task on the target stimulus (i.e., categorizing the target stimulus as a word or nonword). The general semantic priming effect is that word targets are responded to faster when they are semantically related to the preceding prime word, compared with when they are unrelated to the prime (Neely, 1977; see also Goldinger, 1996). In Experiment 2 of Ostrand et al. (2016), some of the primes were audiovisually congruent, meaning that the speaker’s voice and lip movements produced the same word stimulus (e.g., auditory ‘bait’ + visual ‘bait’ or auditory ‘date’ + visual ‘date’), while others were incongruent “McGurk” stimuli (i.e. auditory ‘bait’ + visual ‘date’—often perceived as ‘date’; see also Barutchu et al., 2008; Brancazio, 2004; Sams et al., 1998; see also MacDonald & McGurk, 1978; McGurk & MacDonald, 1976, for background on predicted McGurk effects). The authors found that the incongruent stimuli produced a pattern of semantic priming more similar to the priming found for words that were audio-visually congruent and matched the incongruent *auditory* word, than the priming effect for words that were audio-visually congruent and matched the (ostensibly perceived) incongruent visual word. For instance, the incongruent item formed from auditory ‘bait’ + visual ‘date,’ while putatively perceived as ‘date,’ primed the word semantically related to ‘bait’ (‘worm’) but not the word semantically related to ‘date’ (‘time’).

Ostrand et al. (2016) concluded that, for incongruent multisensory stimuli, initial lexical access and semantic processing operates on the *auditory* unimodal signal rather than the visual unimodal or integrated—and perceived—speech signal. The results suggest that lexical access can initially occur on the auditory signal alone *before* integration of the auditory and visual signals is complete. These results are surprising considering that with the McGurk effect, participants typically report *hearing* the integrated word. That semantic processing seems to operate on auditory-only information even though participants subsequently perceived the integrated word suggests that at least first-pass lexical access occurs prior to the completion of multisensory integration.

The finding that the auditory component of incongruent words is used for lexical access challenges the perspective that audiovisual integration occurs early (e.g., Rosenblum et al., 2016a, b), at the prephonemic feature level of speech perception, and (largely) before lexical access. These findings could indicate that a shift in theories of multisensory speech perception is needed. Indeed, recent work has proposed theories of speech processing that can accommodate these findings (Baart & Samuel, 2015; Mitterer & Reinisch, 2017; Samuel & Lieblich, 2014). For example, Mitterer and Reinisch (2017) propose that diffusion of cognitive resources can impede the use of multisensory information. Thus, unlike traditional theories of multisensory integration, these authors argue that multisensory integration is not automatic. These authors suggest that the lexical decision task adopted by Ostrand et al. (2016) may have imposed too great of a cognitive load for the multisensory information from the primes to be effective.

Separately, Samuel and Lieblich (2014); see also Baart & Samuel, 2015), propose that rather than all speech processes operating on the integrated multisensory information (i.e., Rosenblum et al., 2016a, b), there are two separate cognitive processes involved in speech processing. The first is a perceptual process that relates to how speech is identified, while the second is a linguistic process that relates to the meaning derived from the speech signal. Under this theory, the perceptual, but not the linguistic, process deals with multisensory information. Thus, Samuel and Lieblich (2014) argue, incongruent auditory and visual speech can result in a dissociation between the perception of speech and its lexical processing, accounting for the results reported by Ostrand et al. (2016). This account contrasts with traditional accounts by assuming that speech processing is separate from multisensory speech perception.

Given the theoretical import of the Ostrand et al. (2016) findings, the present work reexamines the question of semantic priming with incongruent stimuli. The degree to which the stimuli in Ostrand et al. (2016) elicited McGurk effect perceptions almost certainly varied between items, and testing the relationship between those perceptions and semantic priming is critical for understanding the results of that study. In the present work, we largely replicate the original experimental design, but change the specific stimuli to items which we expect to have a very high rate of inducing McGurk perceptions. This is an important extension of the original work because if the stimuli in Ostrand et al. (2016) did not always induce the McGurk effect—meaning participants heard the auditory word of the incongruent stimuli, rather than the integrated McGurk percept—then the finding of semantic priming consistent with the auditory channel of those stimuli could be attributed to the lack of perception of the integrated percept. Thus, such a result would not necessarily indicate that lexical access precedes multisensory integration, but rather that which signal is used for lexical access is dependent on how well the two unimodal, incongruent signals are integrated. In the present study, we will be using a different set of McGurk compatible stimuli: auditory 'b' and visual 'v'-initial words. Prior research has demonstrated that the auditory 'ba' + visual 'va' McGurk combination is very reliable for inducing an illusory "heard" response (e.g., 99% 'va' perceptions; Saldaña & Rosenblum, 1994; see also Rosenblum & Saldaña, 1992). Moreover, in the present experiment, the same set of participants performed both the priming task and a free response identification task of the incongruent stimuli to measure the size of the McGurk effect. This was instituted to establish that for the particular tested group of participants, the visual influence was reliable and occurred in the predicted way. Inclusion of a McGurk identification task also allows for assessing the correlation between the degree of semantic priming and the identification of the stimuli.

Experiment 1: Testing the relationship between speech identification and lexical access

In this first experiment, we tested the relationship between the degree of semantic priming induced and the perception/identification of the incongruent prime word. If lexical access of audiovisual stimuli occurs on the auditory component, then the present stimuli should induce priming effects from the auditory word, following the results of Ostrand et al. (2016). In addition, the size of the priming effect should not correspond with the McGurk identification results. In contrast, if lexical access is based on the integrated percept rather than the auditory component of audiovisual words, then the stronger /b–v/ McGurk segments used in the present experiment should induce priming effects based on the visual component—and perceived word—of the primes. Further, the size of the visual/perceptual-based priming effect should correspond to the observed consistency of the perceptual McGurk effect for each stimulus.

Method

The materials, design, and procedure of this experiment followed that of Ostrand et al. (2016) Experiment 2, and are identical except where noted. The main change was the use of different stimuli, whose initial segments (/b/ and /v/) are known to induce a strong and consistent McGurk effect. Additional procedural/design modifications from the original experiment included: fewer stimuli items (24 in the present experiment compared with 36 in the original experiment; necessary due to restricting the initial segments to b/v word pairs); and a smaller sample size (119 participants in the present priming experiment compared with 144 in the original experiment).

Participants

Participants were native English speakers from the University of California, Riverside, and provided informed consent to participate. All procedures were approved by the University of California, Riverside Institutional Review Board. All participants reported having normal hearing and vision. All participants were compensated with either course credit or $10.00 cash. The following experiment consisted of two parts, a priming task and an identification task. For the priming task, 119 people participated. Using the R package ‘pwr’ a power analysis found that this sample size had 92.6% power to detect an effect the same size as was reported by Ostrand et al. (2016). One hundred people who participated in the priming task participated in the identification task. The remaining 19 participants did not complete the identification task.[1]

Materials

The stimuli were audiovisual word primes followed by auditory-only word or nonword targets. A 50 ms interstimulus interval (ISI) separated the offset of the prime and the onset of the target, following Ostrand et al. (2016). All stimuli were produced in a single recording session by a male, monolingual native English speaker. The speaker had lived in Southern California for approximately 4 years prior to recording. The videos showed the talker’s full face, from the crown of the head to the tops of his shoulders.

As in the original Ostrand et al. (2016) paper, our central question concerned the semantic priming induced by incongruent stimuli and which information is used in the process of lexical access—the pre-integration auditory stimulus, or the post-integration audiovisual percept. The major change from the original paper was the specific word stimuli used to create the incongruent (and corresponding congruent) primes and their corresponding targets. The items of this experiment were selected with the goal of increasing the likelihood that participants would experience a visual-dominance McGurk effect and “hear” the visually indicated word. The incongruent primes consisted of pairs of English words differing only in their initial consonant. The critical incongruent stimuli were composed of pairs of words that began with ‘b’ and ‘v’ (e.g., auditory ‘bale’ + visual ‘veil’). Prior research has shown that the auditory ‘b’ + visual ‘v’ combination produces a high rate of visually dominated percepts (e.g., 99%; Saldaña & Rosenblum, 1994; see also Rosenblum & Saldaña, 1992). This type of stimuli should increase the likelihood that participants perceive the visual word (i.e., auditory ‘bale’ + visual ‘veil’ perceived as ‘veil’) and decrease the likelihood that they perceive either the auditory word of the incongruent stimulus (‘bale’) or a fusion of the two unimodal signals (e.g., ‘gale’).

We identified 24 /b/-initial–/v/-initial word minimal pairs to be used as critical incongruent stimuli (see Table 1). A pilot study consisting of 27 participants was conducted to test the strength of the visual influence of these word combinations. Using an open response identification task, it was found that these 24 audio-B + visual-V incongruent words produced visually dominated responses (i.e., participants reported perceiving the /v/-initial word of the minimal pair) 75.9% of the time. While this average is notably smaller than the ‘b/v’ visual dominance reported in a prior study (e.g., 99%; Saldaña & Rosenblum, 1994; see also Rosenblum & Saldaña, 1992), it should be noted that these previous reports tested perception of simple syllable stimuli in two-alternative force-choice tasks and the present study’s open-ended response task can produce a higher degree of variability between participants’ responses, thereby reducing the consistency of the visually dominated perception (see Alsius et al., 2018, for a discussion of the effect of forced choice tasks on McGurk

Table 1 Primes and targets

Prime		Audio associates		Visual associates	
Audio	Visual	Related	Unrelated	Related	Unrelated
Bale	Veil	Hay	Exile	Wedding	Disappear
Ballad	Valid	Song	Sell	True	Want
Ballet	Valet	Dance	Break	Parking	Machine
Ban	Van	Stop	Hay	Car	Wedding
Bane	Vein	Curse	Stop	Blood	Car
Banish	Vanish	Exile	Dig	Disappear	Much
Base	Vase	Bottom	Curse	Flowers	Blood
Bat	Vat	Ball	Bottom	Tub	Flowers
Beer	Veer	Drink	Ball	Swerve	Tub
Bender	Vendor	Fender	Dance	Seller	Parking
Bending	Vending	Break	Drink	Machine	Swerve
Bent	Vent	Broken	Song	Air	True
Best	Vest	Worst	Broken	Clothes	Air
Bet	Vet	Money	Worst	Animals	Clothes
Bigger	Vigor	Smaller	Water	Strength	Elect
Bile	Vial	Stomach	Money	Potion	Animals
Boat	Vote	Water	Stomach	Elect	Potion
Bolt	Volt	Nut	Fender	Shock	Seller
Bow	Vow	Down	Nut	Marriage	Shock
Bowel	Vowel	Movement	Smaller	Letter	Strength
Bowl	Vole	Dish	Down	Mouse	Marriage
Burst	Versed	Explode	Dish	Well	Mouse
Bury	Very	Dig	Movement	Much	Letter
Buy	Vie	Sell	Explode	Want	Well

Table 1 summarizes the critical stimuli of Experiment 1. Column 1 shows the words used in the aud-congruent prime conditions, and that are used for the incongruent auditory stimulus. Column 2 shows the words used in the vis-congruent prime conditions, and that are used for the incongruent visual stimulus. The remaining columns of each row display the target words that are related and unrelated to the words in Columns 1 and 2

---

[1] This attrition includes participants who chose not to continue with the experiment after completing the lexical decision task, computer failures that resulted in the identification task not functioning at the time of testing, and three participants who provided implausible responses to all stimuli and thus no identification data were available to analyze.

rates). Moreover, this result is comparable to what is reported in other open-ended response studies of McGurk words (e.g., ~55%; Brancazio, 2004).

The stimuli selection and counterbalancing followed that of Ostrand et al. (2016; Experiment 2). Across participants, each prime was paired with four targets: a target semantically related to the auditory word, a target unrelated to the auditory word, a target semantically related to the visual word, and target unrelated to the visual word (see the Lexical Decision Task procedure section for more details). The related word targets were selected from the University of South Florida Free Association Norms database (Nelson et al., 1998) and the Edinburgh Associative Thesaurus (Kiss et al., 1972), as well as a norming study conducted on students from a similar participant pool as drawn from for the main experiment at UC Riverside and UC San Diego (N = 124). From these three sources, we chose the targets that optimized semantic relatedness to the prime and, when possible, avoided phonological similarity between primes and targets. This choice was made because prior work has found that visual speech stimuli can phonologically prime audio-only speech targets (e.g., Fort et al., 2013). The complete list of critical prime and target stimuli are given in Table 1.

Each semantically-related target was presented as the unrelated target for another prime of the same modality and thus acted as its own control in terms of lexical properties such as length, frequency, and age of acquisition. For example, a given target word (e.g., *hay*) was used as both a *related* target (for the prime: auditory 'bale' + visual 'veil'), as well as an *unrelated* target (for the prime: auditory 'ban' + visual 'van'). Although different target words differ on multiple dimensions, these differences are fully controlled by the design, as these item-specific properties which could affect lexical decision reaction times contribute to both the Related as well as Unrelated reaction times. Thus, when looking at priming effects—namely, the difference between related and unrelated reaction times—those item-specific effects will be cancelled out. Nonword targets and filler primes (see the lexical decision task procedure section for more details) were selected from those used in the original Ostrand et al. (2016) experiment and were recorded by the same talker and during the same recording session as were the critical primes. All auditory stimuli were presented through sound-insulated headphones at an average of 70 dB. The stimuli used for the lexical decision task are available online (10.17605/OSF.IO/AD52R).

Procedure

The experiment procedure contained two parts. First, participants performed a lexical decision task to measure semantic priming from audio-visually incongruent and congruent word stimuli. Second, participants performed an identification task that assessed their perceptions of the audio-visually incongruent and congruent words that were used as primes in the lexical decision task, as well as the audio-only versions of those stimuli.

Lexical decision task During the lexical decision task, participants were instructed to watch and listen to the audiovisual prime word, and then listen to the audio-only target, and indicate if the target item was a word or nonword by pressing one of the two labeled buttons on a button box. Participants were instructed to respond as quickly and accurately as possible. Each session began with six unscored practice trials before the main experiment. The word/nonword button assignment was counterbalanced across participants.

Each participant received 72 prime-target items in the lexical decision task. Half of the primes (36 items) were paired with word targets, and the other half (36 items) were paired with nonword targets. The 24 critical 'b'/'v'-initial items shown in Table 1 were always paired with word targets. The remaining 48 prime items were included as filler trials, 12 paired with word targets and 36 paired with nonword targets.

The counterbalancing design of the experiment is shown in Fig. 1. For a given participant, one third of the 24 critical 'b'/'v'-initial primes (eight items) were presented as incongruent stimuli (e.g., auditory 'bale' + visual 'veil'). Another one third (eight items) were presented as b-initial audiovisual congruent primes (*aud-congruent*), made up of the auditory and visual signals matching the incongruent auditory component (e.g., auditory 'bale' + visual 'bale'). The final one third (eight items) of the critical items were presented as v-initial audiovisual congruent primes (*vis-congruent*), made up of the auditory and visual signals matching the incongruent visual component (e.g., auditory 'veil' + visual 'veil'). The assignment of a particular item to the *incongruent*, *aud-congruent*, or *vis-congruent* condition was counterbalanced across participants.

The remaining 48 trials were filler items. Filler primes included both congruent and incongruent formats and were not restricted to words with the initial consonants of 'b'/'v' (e.g., congruent: 'tease,' 'hog'; incongruent: auditory 'pug' + visual 'tug' putatively perceived as 'tug'; auditory 'might' + visual 'night' putatively perceived as 'night'; see Appendix 1 for the complete list of filler primes). In a departure from the experimental design of the Ostrand et al. (2016) experiment, some filler primes were paired with word targets in addition to nonword targets. Twelve filler primes, divided between incongruent and congruent formats, were paired with word targets. This was done to reduce the potential for participants' learning that only 'b'- and 'v'-initial words preceded word targets, and all other

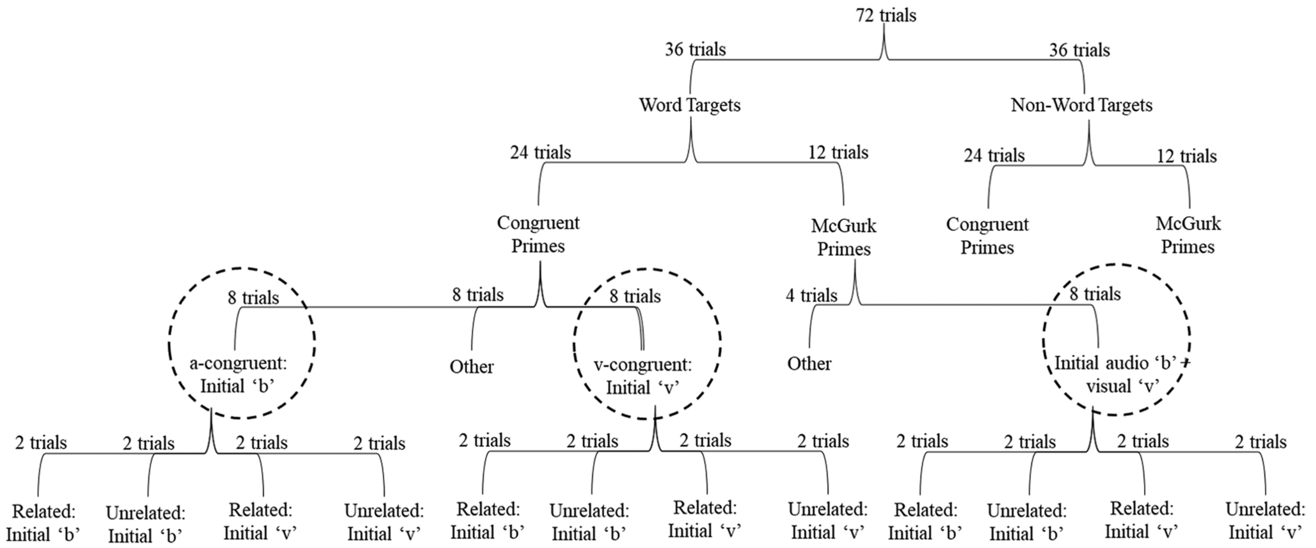

Fig. 1 Trial counts of each stimulus type. *Note.* Illustrates the counterbalancing of conditions in the lexical decision task of the main experiment. The critical trials, circled above, had either aud-congruent, vis-congruent, or incongruent primes. Each participant received 24 of these critical primes, equally divided across the conditions. All other trials were filler and/or nonword target trials. The bottom row of this chart shows the types of targets that followed each of the critical primes

initial phonemes preceded nonword targets. The remaining filler primes (36 items) were paired with nonword targets. Of these, the primes were again divided between incongruent primes (12 trials) and congruent primes (24 trials). Unlike the critical trials in which the specific items that were presented as congruent or incongruent were counterbalanced across participants, every participant received the same filler primes.

Each critical prime stimulus was paired with four types of targets: targets (1) *related* and (2) *unrelated* to the incongruent visual word, and targets (3) *related* and (4) *unrelated* to the incongruent auditory word. Unrelated targets were formed by taking the related targets and re-assigning them to different primes, within modality. For example, ‘song’ was presented as the semantically related target for ‘ballad’ for some participants; and for other participants, it was reassigned to a different prime item to become the unrelated target for ‘bent.’ These four target conditions were crossed with the three prime conditions (incongruent [e.g., audio ‘ballad’ + visual ‘valid’], aud-congruent [audio-visually congruent with incongruent auditory word; e.g., audio ‘ballad’ + visual ‘ballad’], vis-congruent [audio-visually congruent with incongruent visual word; e.g., audio ‘valid’ + visual ‘valid’]), resulting in 12 conditions across the 24 critical items. Thus, a given participant received two items in each critical prime-target condition (note that this is a departure from the original experiment, in which, having 36 critical items, participants received three items in each critical prime-target condition). Each critical prime was only presented once to each participant, and which primes were placed in which condition was counterbalanced across participants.

All subjects were run in a sound-insulated lab room in front of a computer screen that presented all visual stimuli. Participants listened to auditory stimuli using Sony MDR 7506 headphones with volume set to a comfortable listening level. Participants were seated approximately 30 inches from the computer screen. Each trial began with a blank black screen for 1,400 ms followed by a white ‘*’ fixation point presented for 600 ms, resulting in an effective intertrial interval of 2,000 ms. Immediately following the fixation point, the face of the talker appeared and spoke the prime word (the fixation point was aligned with the center of the talker’s lips). After the articulation of the prime word, the screen went blank. Following the procedure in Ostrand et al. (2016), 50 ms after the acoustic offset of the prime word, the audio-only target stimulus was presented, without any accompanying visual stimulus on the screen. The trial ended when the participant pressed either the ‘Word’ or ‘Nonword’ button on a button box. The word/nonword button assignment was counterbalanced across participants. Participants were given one short break administered half-way through the session (between trials 36 and 37).

Perceptual identification task Following the completion of the lexical decision task, participants started the identification task for the critical prime items. Participants were presented with a series of audiovisual prime stimuli and audio-only versions of those primes from the lexical decision task, and were instructed to attend to each utterance and use the keyboard to type the word they *heard* the talker say (following Brancazio, 2004). During this task, participants responded to each critical incongruent prime stimulus (24 items) and each corresponding audiovisual congruent stimulus (48 items), as well as audio-alone versions of the audiovisual congruent items (48 items). Participants were presented with each item twice (for a total of 240 trials) and thus they provided two perceptual identifications for the same stimulus. For each item for each participant, responses were tabulated for proportion of responses with initial consonant consistent with the incongruent auditory (i.e., initial 'b') and the proportion of responses consistent with the incongruent visual (i.e., initial 'v') for all critical items (i.e., incongruent, aud-congruent, and vis-congruent).

A programing error resulted in one of the 24 incongruent items (audio 'buy' + visual 'vie') being presented to only 15 participants, rather than the full set of 100 participants.[2] Given the reduced sample size for this item relative to the rest of the set, this item was excluded from all identification task analyses (though this item was retained in analyses of the priming task, with the exception of the ANCOVA and associated correlation as there was no covariate data from the identification responses). Thus, in the identification task, most participants were presented with 23 incongruent critical items along with the corresponding 48 audiovisual congruent items that were used as primes for the lexical decision task. In addition, they were also presented with the 48 audio-alone versions of the audiovisual congruent items.

Stimuli were presented in random order. Participants were not informed that the stimuli were the same items from the lexical decision task, and were not informed that the items would all be words. Participants were allowed to view their responses as they typed them and were instructed to correct any errors or typographic mistakes before proceeding to the next trial. As in the priming task, each audiovisual trial included a fixation point at the location of the talker's lips that was present for 600 ms immediately preceding the appearance of the talker's face (or the time when the talker's face would appear for the audio-only trials). The perceptual identification task took approximately 15 minutes.

Results

Semantic priming reaction times

The data used for this analysis are available online (10.17605/OSF.IO/AD52R). The analysis strategy (including criteria for trial exclusions) followed the same procedure used in Ostrand et al. (2016), Experiment 2. Reaction times were measured from target offset. Only reaction times from trials that included one of the 24 critical primes (the McGurk words) and their 48 congruent counterparts were analyzed (i.e., filler prime items, and those with nonword targets, were not included in the analysis). Responses that occurred before the target word onset (0.6%), were incorrect (7.5%), or that were more than two standard deviations from the condition mean reaction times (6.0%) were excluded from the analysis. These criteria meant that 25 participants contributed no data to at least one condition of the analysis and thus were excluded from the analysis of participant responses (F_1). This relatively high participant exclusion rate is likely the result of the restricted set of possible stimuli—since each participant received only two items in each condition, across 12 conditions, it was easy for a participant to have both items in at least one condition excluded for the reasons mentioned above. Although this participant exclusion rate is higher than anticipated, it was for this reason that a relatively large participant sample was tested in the first place. It is, however, worth noting, that the trial-level exclusion rates are similar to those reported in Ostrand et al. (2016); it is the reduced number of stimuli per participant in the current experiment that led to a much higher participant-level exclusion rate.

Reaction times were submitted to both a participant (F_1) and an item (F_2) analysis. Each analysis began with an omnibus analysis of variance (ANOVA) consisting of the following factors: 2 relatedness (related vs. unrelated) × 2 target (associated with: visual word vs. auditory word) × 3 prime (incongruent, vis-congruent, or aud-congruent). Condition means for the participant analysis are displayed in Fig. 2.

Consistent with the results of Ostrand et al. (2016) there was no main effect of prime, $F_1(2, 186) = 0.20$, $p = .819$, $\eta_p^2 < .01$; $F_2(2, 46) = 0.18$, $p = .834$, $\eta_p^2 = .01$; M_1: incongruent: 329 ms, vis-congruent: 333 ms, aud-congruent: 335 ms. As was also found by Ostrand et al. (2016), there was a significant main effect of relatedness, $F_1(1, 93) = 28.16$, $p < .001$, $\eta_p^2 = .23$; $F_2(1, 23) = 4.62$, $p = .042$, $\eta_p^2 = .17$, indicating that across conditions, targets were identified as words faster when they were semantically related to the preceding prime than when they were unrelated (M_1: related: 310 ms vs. unrelated: 354 ms). Finally, the participant, but not the item, analysis showed a significant main effect of target, $F_1(1, 93) = 37.19$, $p < .001$, $\eta_p^2 = .29$; $F_2(1, 23) = 4.18$, $p = .052$, $\eta_p^2 = .15$. The effect of Target in the by-participants analysis

---

[2] An artifact of this error was that the presentation of the other items during the identification task was not balanced across subjects. While most items were presented twice to each participant during the identification task, a random subset of items (6–-8 items, including incongruent, congruent, and audio-only items) was presented three times to individual participants. Additionally, the majority of participants received the incongruent item audio 'bane' + visual 'vein' and audio-only 'vein' 3 times, with the remainder receiving it two or four times. As the analyses of the identification data are based on the within-item means, not the individual trial responses, this error should not have a major effect on the pattern of results.

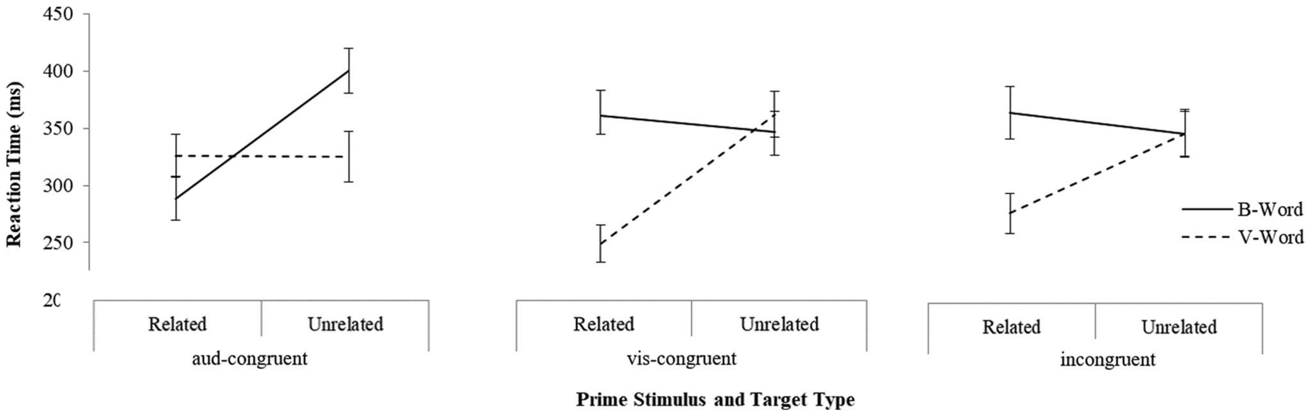

Fig. 2 Reaction time to targets related and unrelated to different prime types. *Note.* The values on the vertical axis are reaction times following target offset. Solid lines correspond to targets related or unrelated the incongruent auditory word. Broken lines correspond to targets related or unrelated to the incongruent visual word. Error bars show the standard error of the mean. Data tabulated by participants (F_1 analysis)

indicates that both related and unrelated targets associated with the visual word (e.g., ‘veil’➔‘wedding’) were identified faster than targets both related and unrelated to the auditory word (e.g., ‘bale’➔‘hay’; M_1: visual: 314 ms vs. auditory: 351 ms). None of the two-way interactions in the omnibus test were significant in either the participant or item analyses.

The most important effect returned by the omnibus test is the three-way interaction between relatedness, target association, and prime stimulus. Both the participant, $F_1(2, 186) = 19.99$, $p < .001$, $\eta_p^2 = .18$, and the item, $F_2(2, 46) = 20.86$, $p < .001$, $\eta_p^2 = .48$, analyses revealed that this interaction was significant. This interaction is portrayed in Fig. 2. This interaction indicates that whether a prime produced semantic priming to the auditory-associated targets or the visual-associated targets depended on whether the prime was an incongruent stimulus, vis-congruent stimulus (consistent with the incongruent’s visual component), or aud-congruent stimulus (consistent with the incongruent’s auditory component). Importantly, it was this three-way interaction, and subsequent analyses, that allowed Ostrand et al. (2016) to conclude that the incongruent prime induced priming responses more similar to the auditory than visual component of the incongruent stimulus.

However, the pattern of results portrayed in Fig. 2 tells a different story from that of the previous paper. Numerically, targets associated with the *visual* channel show a greater priming effect (i.e., related targets were responded to faster than unrelated targets) compared with targets associated with the auditory channel, for the incongruent (auditory target: −17 ms [unrelated – related]; visual target: 67 ms [unrelated – related]) and vis-congruent primes (auditory target: −15 ms [unrelated – related]; visual target: 116 ms [unrelated – related]). In contrast, when the prime was aud-congruent (and consistent with the audio component of the incongruent stimulus), targets associated with the auditory channel showed a numerically greater priming effect than did targets associated with the visual channel (auditory target: 111 ms [unrelated – related]; visual target: −1 ms [unrelated – related]). Thus, as can be seen in Fig. 2, priming responses to the incongruent stimulus appear to be more similar to the vis-congruent than aud-congruent stimuli.

As the Prime factor had three levels, additional analyses were needed to verify these numerical patterns and determine the true locus of the interaction, and whether it indicates that the effect was driven by the difference between the aud-congruent condition (audio-visually congruent with the incongruent *auditory word*) relative to the incongruent and vis-congruent conditions, as suggested by the numerical results.

Planned comparisons

To identify the locus of the interaction, we computed ANOVAs examining each pairing of 2 of the 3 prime conditions in 2 (relatedness) × 2 (target) × 2 (prime) ANOVAs. Again, these analyses were computed by participants (F_1) and by items (F_2). The results of these analyses are shown in Table 2. The most important results of these analyses are the three-way interactions that indicate that the priming effect for auditory-associated and visual-associated targets is modulated by the prime condition. As can be seen in Table 2, this three-way interaction is present when comparing the incongruent and aud-congruent primes and when comparing the aud-congruent and vis-congruent primes. In contrast, this interaction was not significant when comparing the incongruent and vis-congruent primes, suggesting that the pattern of priming is most similar between the incongruent and vis-congruent prime items.

Table 2 Results of post hoc analysis of Experiment 1

		By participants	
Factor	Incongruent vs. aud-congruent	Incongruent vs. vis-congruent	Aud-congruent vs. vis-congruent
Prime	$F(1, 93) = 0.40, p = .531, \eta_p^2 < .01$	$F(1, 93) = 0.18, p = .671, \eta_p^2 < .01$	$F(1, 93) = 0.04, p = .843, \eta_p^2 < .01$
Target	$F(1, 93) = 11.77, p = .001, \eta_p^2 = .11$*	$F(1, 93) = 24.03, p < .001, \eta_p^2 = .21$*	$F(1, 93) = 20.02, p < .001, \eta_p^2 = .18$*
Relatedness	$F(1, 93) = 14.38, p < .001, \eta_p^2 = .13$*	$F(1, 93) = 13.23, p < .001, \eta_p^2 = .13$*	$F(1, 93) = 29.15, p < .001, \eta_p^2 = .24$*
Prime × Target	$F(1, 93) = 1.09, p = .299, \eta_p^2 = .01$	$F(1, 93) = 0.14, p = .711, \eta_p^2 < .01$	$F(1, 93) = 1.88, p = .174, \eta_p^2 = .02$
Prime × Relatedness	$F(1, 93) = 2.68, p = .105, \eta_p^2 = .03$	$F(1, 93) = 1.28, p = .261, \eta_p^2 = .01$	$F(1, 93) = 0.05, p = .819, \eta_p^2 < .01$
Target × Relatedness	$F(1, 93) = 0.48, p = .491, \eta_p^2 = .01$	$F(1, 93) = 22.49, p < .001, \eta_p^2 = .20$*	$F(1, 93) = 0.18, p = .675, \eta_p^2 < .01$
3 way	$F(1, 93) = 18.94, p < .001, \eta_p^2 = .17$*	$F(1, 93) = 1.61, p = .208, \eta_p^2 = .02$	$F(1, 93) = 36.35, p < .001, \eta_p^2 = .28$*
		By Items	
Factor	Incongruent vs. aud-congruent	Incongruent vs. vis-congruent	Aud-congruent vs. vis-congruent
Prime	$F(1, 23) = 0.47, p = .500, \eta_p^2 = .02$	$F(1, 23) = 0.13, p = .720, \eta_p^2 = .01$	$F(1, 23) = 0.04, p = .844, \eta_p^2 < .01$
Target	$F(1, 23) = 2.56, p = .123, \eta_p^2 = .10$	$F(1, 23) = 4.89, p = .037, \eta_p^2 = .18$	$F(1, 23) = 3.47, p = .075, \eta_p^2 = .13$
Relatedness	$\mathrm{F}(1, 23) = 2.78, p = .109, \eta_p^2 = .11$	$F(1, 23) = 4.29, p = .050, \eta_p^2 = .16$	$F(1, 23) = 6.34, p = .019, \eta_p^2 = .22$
Prime × Target	$F(1, 23) = 0.61, p = .444, \eta_p^2 = .03$	$F(1, 23) = 0.34, p = .563, \eta_p^2 = .02$	$F(1, 23) = 1.72, p = .203, \eta_p^2 = .07$
Prime × Relatedness	$F(1, 23) = 1.67, p = .209, \eta_p^2 = .07$	$F(1, 23) = 5.21, p = .032, \eta_p^2 = .19$	$F(1, 23) = 0.19, p = .669, \eta_p^2 = .01$
Target × Relatedness	$F(1, 23) = 0.09, p = .763, \eta_p^2 < .01$	$F(1, 23) = 18.57, p < .001, \eta_p^2 = .45$*	$F(1, 23) = 0.82, p = .374, \eta_p^2 = .03$
3 way	$F(1, 23) = 18.26, p < .001, \eta_p^2 = .44$*	$F(1, 23) = 3.27, p = .084, \eta_p^2 = .13$	$F(1, 23) = 46.07, p < .001, \eta_p^2 = .67$*

The top panel to Table 2 shows the results from the F_1 analyses examining two levels of the prime condition, the bottom panel displays the results from the F_2 analyses. In both panels, the first column shows results when the prime factor included incongruent primes and aud-congruent primes. The second column shows the results when the prime factor included the incongruent primes and vis-congruent primes. The third column shows the results when prime included the two audiovisual-congruent conditions. Asterisks indicate results that were statistically significant at $\alpha = .015$ (i.e. Bonferroni corrected for three post hoc tests). The critical result is the three-way interaction shown in the last row of each panel and is significant for the incongruent versus aud-congruent and aud-congruent versus vis-congruent but not the incongruent versus vis-congruent columns for both panels

Together these results indicate the priming effect on auditory-associated and visual-associated targets is modulated by prime stimulus type (aud-congruent, vis-congruent, incongruent). The incongruent and vis-congruent primes induce similar patterns of priming as each other, both of which are different from those induced by aud-congruent primes. This suggests that semantic priming from the incongruent primes was consistent with priming from the visual, rather than auditory, signal. These results contrast with the results reported by Ostrand et al. (2016), who found that it was their incongruent and *audio*-congruent primes which induced similar responses. One possible reason for this difference from the Ostrand et al. (2016) work could be differences in the rate at which participants perceived incongruent stimuli as matching the visual signal, or as matching the auditory signal. This possibility will be explored in the subsequent section focused on the identification results for the stimuli of this experiment, as well as in a post-hoc experiment detailed below.

Identification task responses

Next, we tested whether the semantic priming results corresponded to the identification of the incongruent primes. After the lexical decision task, participants performed a perceptual identification task in which they wrote what word they heard for each stimulus. In analyzing these identification responses, we had to consider how best to measure the McGurk effect. The operational definition of the McGurk effect varies in the literature, with some researchers defining only identifications that differ from both the auditory *and* visual stimulus as the McGurk effect (e.g., Magnotti & Beauchamp, 2015; van Wassenhove et al., 2007) while others define the effect as any instance in which the visual stimulus changes the perception of the auditory stimulus (e.g., Rosenblum & Saldaña, 1992; see also Alsius et al., 2018). However, in the present work, we are concerned with the relative consistency with which prime items were identified as the particular words that made up the auditory or visual signal (as the target stimuli were related/unrelated to one or the other of those words). Therefore, we calculated two identification rates for each incongruent item: the percentage of auditory word responses and the percentage of visual word responses. Note that because we used an open-response task, participants could provide responses that corresponded to neither the auditory nor visual word and thus the sum of auditory and visual identifications was not necessarily 100% of responses for each incongruent stimulus. Using these two measures of the identification of the incongruent stimulus enables us to see the proportion of participants' perception of the incongruent stimulus as the visual signal as compared with the auditory signal. This is important because the lexical

Table 3 Identification rates by item for different stimulus types

	Incongruent		Audiovisual-congruent		Audio-only	
	Initial consonant of identification response		Stimulus initial consonant (% 'V' responses)			
Auditory Word	% 'V'	% 'B'	V-Word	B-Word	V-Word	B-Word
Bale	78.1	4.6	96.5	29.8	95.5	66.0
Ballad	77.6	20.4	100.0	6.9	99.0	22.7
Ballet	70.9	28.1	96.6	9.7	97.0	14.4
Ban	85.5	8.0	97.0	7.1	96.9	4.5
Bane	78.5	12.6	92.5	26.0	82.4	45.4
Banish	70.9	27.1	98.0	16.4	98.4	27.6
Base	75.5	8.0	96.1	10.7	96.5	53.3
Bat	63.7	8.5	76.2	13.4	59.0	33.7
Beer	19.6	10.6	16.0	1.0	16.5	2.6
Bender	63.5	28.4	87.6	10.6	87.5	7.7
Bending	75.4	19.1	98.5	6.1	97.5	14.7
Bent	79.0	17.5	99.0	6.5	99.0	32.8
Best	81.5	13.8	96.0	2.0	90.9	27.2
Bet	82.1	13.8	97.0	11.8	97.4	21.8
Bigger	58.3	30.2	82.2	16.8	83.4	11.6
Bile	76.8	9.6	99.0	19.6	99.0	31.0
Boat	64.7	33.3	98.5	11.7	99.0	40.0
Bolt	80.3	15.8	99.0	26.9	99.0	35.2
Bow	75.7	22.8	89.4	5.6	77.8	11.3
Bowel	87.0	11.5	99.0	19.2	94.1	28.9
Bowl	53.5	31.8	91.0	8.6	85.3	10.7
Burst	25.1	24.2	82.9	5.6	86.9	11.2
Bury	48.5	50.0	83.6	12.8	89.0	13.3

This table shows the mean identifications of each item. For incongruent stimuli, the percentage of identification responses which had an initial 'b' (matching the auditory channel) and an initial 'v' (matching the visual channel) are presented. For the audiovisual-congruent and audio-only stimuli, the percentage of identification responses which had an initial 'v' are presented, because erroneous 'v' identifications could inflate the observed McGurk effect for a given stimulus item

decision task results hinge on which signal of the incongruent stimulus is used for lexical access, and the relationship between the perception of the stimuli and lexical access. This point is further discussed below in the "Preparing McGurk Identifications For Semantic Priming Analysis" section.

McGurk identifications

We analyzed our data by tabulating participant responses that began with the letter 'b' and those that began with the letter 'v.' One hundred of the participants who completed the lexical decision task subsequently completed the identification task. A further two participants had incomplete data files due to problems with the experiment program and were thus excluded from the identification task analyses. For each participant we tabulated the proportion of b-initial and v-initial responses for each incongruent item and averaged these proportions across participants to calculate item means (Table 3).

Our stimuli produced McGurk effects with visually-consistent responses of 68.4% ($SE = 3.6\%$) and auditory-consistent responses of 19.5% ($SE = 2.3\%$). The rate of visually-consistent responses was significantly greater than the rate of auditory consistent responses, $t(22) = 9.87$, $p < .001$, $d = 2.06$, indicating that these stimuli supported robust McGurk effects. The rate of visually-consistent responses was also significantly greater than the rate of erroneous v-initial responses found for the aud-congruent ($M = 12.4\%$, $SE = 1.6$), $t(22) = 16.28$, $p < .001$, $d = 3.39$, and corresponding auditory-only items ($M = 24.7$, $SE = 3.4$), $t(22) = 11.77$, $p < .001$, $d = 2.45$, further supporting the contention that the identification of the incongruent stimuli was associated with multisensory integration. These data are similar to those found for our pilot study and the strength of the effect is comparable to other studies that have used word stimuli and free-response tasks (e.g., Brancazio, 2004; ~55%). Consistent with other research (e.g., Barutchu et al., 2008; Basu Mallick et al., 2015; Brown et al., 2018) there was a wide range in the proportion of visually consistent responses across the different items (ranging from 87.0% for 'vowel' identifications of audio-'bowel' + visual-'vowel' to 19.6% for 'veer' identifications of audio-'beer' + visual-'veer') as well as a wide range in the proportion of auditory-consistent responses (ranging from 50.0% for 'bury' identifications of audio-'bury' + visual-'very' to 4.6% for 'bale' identifications of audio-'bale' + visual-'veil'). Of note, only one item (audio-'bury' + visual-'very') failed to support more visual-consistent than auditory-consistent identifications (Table 3). In other words, 95.7% of items produced more visual-consistent identifications than auditory-consistent identifications. A number of factors likely account for the differences in effect strength across items including visibility of articulation, as well as relative word frequency of the two unimodal signals (e.g., see Dorsi, 2019; see also Barutchu et al., 2008; Brancazio, 2004).

Table 3 also presents the identification rates for audiovisual-congruent stimuli, as well as audio-only stimuli which correspond to each of the two unimodal signals for each incongruent stimulus. The identification data for the congruent and audio-only stimuli is included to serve as a baseline for how these items are perceived when there is no audio-visual conflict. If, for example, the audiovisual *congruent* stimulus auditory 'bane' + visual 'bane' is perceived as "vein" a high percentage of the time, that would suggest that even if the corresponding *incongruent* stimulus auditory 'bane' + visual 'vein' is often perceived as "vein," it is likely due to properties of the audio signal, rather than participants successfully integrating the multimodal signals and perceiving the McGurk effect.

The first row of the table presents identification data for the incongruent stimulus auditory 'bale' + visual 'veil,' and what percentage of participants identified it as "veil" (%V) as compared with "bale" (%B). The next column shows the identification data for the two corresponding audiovisual stimuli—first the vis-congruent (auditory 'veil' + visual 'veil') and then the aud-congruent (auditory 'bale' + visual 'bale')—as the percentage of participant responses that identified each congruent stimulus as the visual component of the corresponding incongruent stimulus (here, 'veil'). The final column shows the identification data for the two corresponding audio-only stimuli—first matching the visual signal of the incongruent item ('veil') and next matching the auditory signal of the incongruent item ('bale'). As with the congruent stimuli, responses are presented as the percentage of participant responses that identified the stimulus as beginning with "V," to give a baseline for how often the audio-alone stimulus—with no influence of a visual signal to induce a McGurk effect—is perceived as the "V" word. To the extent that the incongruent stimuli are identified as the "V" word more often than the corresponding congruent and audio-only stimuli are, is the extent to which the incongruent stimuli successfully induce the McGurk effect in participants.

Preparing McGurk identifications for semantic priming analysis To address our hypothesis that the strength of the visual influence on speech perception modulates the semantic priming by McGurk words, we included the McGurk identification rates as a covariate in the item analysis of the priming reaction time data. For this analysis, we converted the identification responses into McGurk *identification differentials* by subtracting the auditory-consistent response rate from the visually consistent response rate for each incongruent prime. For example, the incongruent stimulus auditory 'bane' + visual 'vein' was perceived as 'bane' (McGurk-auditory) 12.6% of the time and as 'vein' (McGurk-visual) 78.5% of the time. This item thus has a McGurk identification differential of 65.9%, meaning that participants perceived this incongruent item as its visual signal ('vein') 65.9 percentage points more often than they perceived it as its auditory signal ('bane'). In this way, the identification differential conveys the relative frequency of the two perceptions that could be expected to influence semantic priming. Note that this score also has the benefit of excluding nonvisual and nonauditory McGurk responses (e.g., if a participant reported perceiving "lane" or "cane") for which the experimental design was not equipped to assess semantic priming.

Importantly, for each incongruent item, the identification differential was calculated based only on identification responses from the participants who also provided lexical decision task reaction times for that particular incongruent item. Recall that during the semantic priming task, each participant was presented only *eight* critical incongruent primes. Thus, the identification-differential score for each item only included identification data from the specific participants who had been presented that word in incongruent format during the semantic-priming task. Recall also that the reaction times submitted to the semantic priming analyses were subject to exclusion criteria (see first paragraph of Results section). Thus, if a participant's reaction time value for an incongruent stimulus in the priming experiment was excluded from the analysis, their corresponding identification response for that stimulus was also excluded from the identification-differentials calculation. Thus, the number of participants that contributed to each identification-differential ranged between 31 and 37 for each incongruent item.

Interaction of word identification and reaction times

To compute the degree of semantic priming from our design, we needed to compare the reaction times from the twelve conditions. As the participant analysis (F_1) averages reaction

Table 4 Results of the analysis of covariance

Effects	Results
Prime	$F(2, 42) = 3.82, p = .030, \eta_p^2 = .15*$
Prime × Identification-Differential	$F(2, 42) = 4.13, p = .023, \eta_p^2 = .16*$
Target	$F(1, 21) = 1.61, p = .218, \eta_p^2 = .07$
Target × Identification-Differential	$F(1, 21) = 8.55, p = .008, \eta_p^2 = .29*$
Related	$F(1, 21) = 0.92, p = .349, \eta_p^2 = .04$
Related × Identification-Differential	$F(1, 21) = 0.03, p = .873, \eta_p^2 < .01$
Prime × Target	$F(2, 42) = 0.97, p = .388, \eta_p^2 = .04$
Prime × Target × Identification-Differential	$F(2, 42) = 0.37, p = .69, \eta_p^2 = .02$
Prime × Related	$F(2, 42) = 0.11, p = .894, \eta_p^2 = .01$
Prime × Related × Identification-Differential	$F(2, 42) = 0.25, p = .779, \eta_p^2 = .01$
Target × Related	$F(1, 21) = 0.45, p = .511, \eta_p^2 = .02$
Target × Related × Identification-Differential	$F(1, 21) < 0.01, p = .956, \eta_p^2 < .01$
Prime × Target × Related	$F(2, 42) = 8.34, p = .001, \eta_p^2 = .28*$
Prime x Target × Related × Identification-Differential	**$F(2, 42) = 4.31, p = .020, \eta_p^2 = .17*$**

Results of the Target (V word vs. B word associates) × Related (related vs. unrelated) × Prime (incongruent vs. vis-congruent vs. aud-congruent) × Identification-Differential ANCOVA. The key four-way interaction is shown in boldface. Significant effects ($p < .05$) are indicated by asterisks

times across individual items, it is difficult to use this analysis to examine the effect of item identification on reaction times. Instead we included the identification-differential for each item and calculated the item analysis (F_2). Thus, reaction times from the lexical decision task and McGurk identification differential from the identification task were entered into a four-way (Target x Relatedness × Prime × Identification Differential) analysis of covariance (ANCOVA) by items (F_2). This ANCOVA retained the significant three-way interaction between relatedness, target, and prime, $F_2(2, 42) = 8.34$, $p = .001$, $\eta_p^2 = .28$, indicating that the pattern of semantic priming depended on the relationship between prime stimulus type and the prime and target. More importantly, this analysis also returned a four-way interaction between those factors and the identification differential, $F_2(2, 42) = 4.31$, $p = .020$, $\eta_p^2 = .17$. This four-way interaction suggests that the degree of semantic priming was modulated by the strength of the McGurk effect, and is further discussed in the following section. The full results of the ANCOVA are presented in Table 4.

Localizing the correspondence between semantic priming and perception The four-way interaction of the ANCOVA indicates that the relationship between prime, target, and relatedness is dependent on the perception (i.e., the identification differential) of the incongruent priming item. A number of different data patterns could produce this significant interaction, but only one such pattern would support our hypothesis that incongruent primes correspond to semantic priming consistent with the identification of those primes. This section examines this interaction to determine the locus of the effect and determine if the predicted pattern was present (i.e., does semantic priming from the auditory vs. visual channel of the incongruent prime correspond with auditory vs. visual consistent identifications of the incongruent prime).

To do so, priming scores and McGurk identification-differentials were entered into a correlation. Recall that the four-way interaction of the ANCOVA was not driven by any single set of reaction times, but from the relationship across twelve sets of reaction times interacting with the McGurk identification-differential. Just as the McGurk identification-differential needed to be measured in a way that conveyed both the rate of auditory- and visual-consistent identifications, to understand how the strength of the McGurk effect interacted with degree of semantic priming, we need to measure semantic priming in a way that captures the degree of priming from both the auditory and visual components of an incongruent prime.

To this end, we calculated *priming-differential* scores from the lexical decision task reaction times. These priming-differential scores were calculated on the four types of targets following incongruent primes (i.e., targets related and unrelated to the visual and auditory channels of the incongruent prime). From these, the priming-differential scores for each incongruent item were calculated in three steps. (1) For each incongruent word prime, the mean reaction time for the target related to the incongruent auditory word was subtracted from the mean reaction time for the target unrelated to the incongruent auditory word. In this way positive values indicate that reaction times to targets related to the incongruent auditory prime were shorter than the reaction times to targets unrelated the incongruent auditory prime, and therefore indicate that the auditory component of an incongruent stimulus induced semantic priming. (2) This process was repeated for reaction times to targets related and unrelated to the incongruent visual words. (3) For each incongruent stimulus, the auditory priming score (the result of Step 1) was subtracted from the visual priming score (the result of Step 2), forming the *priming differential.*[3] The priming differential indicates how much each incongruent prime induced priming by its visual component relative to priming by its auditory component. A positive priming-differential score would indicate that this incongruent item showed a stronger priming effect from its visual signal than from its audio signal; a negative priming-differential score would indicate that this incongruent item showed a stronger priming effect from its auditory signal than its visual.

The correlation between the *identification-differential* and *priming-differential* scores for each incongruent item was r = .37, p = .042 (one-tailed[4]) for the 23 incongruent items tested[5] (see Fig. 3). This correlation illustrates that the four-way interaction found for the ANCOVA was driven by a relationship in which items that produced stronger McGurk-visual perceptions also produced larger semantic priming effects consistent with the incongruent (visual) signal. This finding is consistent with the hypothesis that semantic priming is related to the *perception* (identification) of the prime.

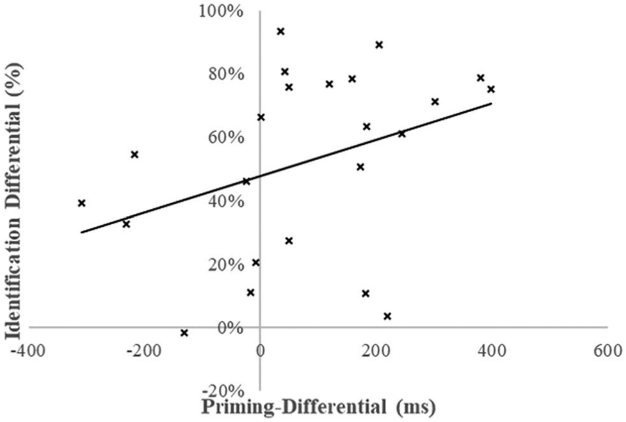

Fig. 3 Priming and prime identification. *Note.* Relationship between semantic priming and the McGurk effect. The vertical axis shows the identification rate of McGurk-visual responses minus the rate of McGurk-auditory responses for each item (the *identification differential*). The identification-differential values only include responses from participants who responded to targets following that incongruent prime in the lexical decision priming task as well. The *priming differential* is shown along the horizontal axis and was calculated by subtracting the difference of reaction times between targets unrelated and related to the incongruent auditory word from the difference of reaction times between targets unrelated and related to the incongruent visual word

Discussion

Taken together, these results suggest that, at least in some contexts, semantic priming can be more consistent with audiovisual word identification than auditory word information. Namely, we find semantic priming consistent with the visual word for incongruent primes that are generally identified as the visual word. Importantly, this is not a simple visual-dominance effect; the degree to which an incongruent prime supports priming to its visual word corresponds to the consistency with which that prime is identified as its visual word. Finally, it should be noted that this conclusion does not preclude the possibility that lexical access, and thus semantic priming, may sometimes operate on the auditory stimulus. We argue, instead, that semantic priming may be generated by the auditory stimulus when the auditory stimulus is what is perceived.

Post hoc experiment: Evaluating word identification from the stimuli of Ostrand et al. (2016)

The question naturally arises as to why the results reported for Experiment 1 differ so drastically from the results reported by Ostrand et al. (2016). Experiment 1 found that (a) semantic priming from the incongruent primes was consistent with the visual word of those primes; (b) those primes reliably produced McGurk effects consistent with the visual channel (i.e., participants *heard* the visual word) and; (c) across items, the degree of semantic priming correlated with the success of the McGurk effect. That is, unlike the findings of Ostrand et al. (2016), Experiment 1 found that semantic priming was consistent with the identification of the incongruent stimulus. One possible

---

[3] The formula for the priming-differential score can thus be summarized as:

priming-differential = [(visual-unrelated – visual-related) – (auditory-unrelated – auditory-related)]

[4] This was a post hoc test to determine whether the relationship between prime identification and semantic priming, which was confirmed by the ANCOVA, occurred in the direction predicted by our hypothesis. We chose to use a one-tailed test a priori, based on the strongly directional hypothesis that semantic priming would be consistent with the identification of the prime stimuli. A *negative* correlation would indicate that semantic priming linearly corresponded to the word that was *not* identified which is not predicted by a competing hypothesis and would be as incompatible with the tested hypothesis as an absence of any correlation (see Cho & Abe, 2013; Kimmel, 1957; Ruxton & Neuhäuser, 2010, for a discussion).

[5] As another method for investigating this relationship, a linear mixed-effects model was conducted with random intercepts for item (prime) and participant. The independent variables were target association (coded as visual associated = −1, auditory associated = 1), relatedness (coded as related = 1, unrelated = −1), and identification-differential. The dependent variable was reaction time to the target word. It found that reaction times for the incongruent primes were significantly predicted by the interaction between Relatedness (related vs. unrelated), target association (associated with auditory vs. visual channel of prime), and McGurk differential; $\hat{\beta} = 19.53$, $SE = 11.00$, $t = 1.78$, $p = .038$ (one-tailed).

explanation for this difference is that the stimuli of these two studies supported different McGurk effects. That is, it is possible that the stimuli of Ostrand et al. (2016) were not identified as the visual word as much as were the stimuli of Experiment 1. This question is examined in this post hoc experiment.

To address this question, we turn to a follow-up experiment which measured McGurk identification rates of the stimuli in the original Ostrand et al. (2016) study, drawn from the same participant population of undergraduate UC San Diego students. These McGurk identification responses were used for this analysis. These previously unreported data are relevant to our question concerning the relationship between the perception of incongruent stimuli and the semantic priming induced by those stimuli.

Method

Participants

Two hundred and eighty-eight students from the University of California, San Diego, participated in this experiment; 33 were excluded for reasons unrelated to the current analysis (e.g., missing reaction time data). The remaining 265 participants were included in the analysis below. All participants were native English speakers and reported having normal hearing and vision, and none had participated in Ostrand et al. (2016) Experiment 2. Participants provided informed consent to participate and were compensated with course credit. All procedures were approved by the University of California, San Diego, Institutional Review Board.

Materials

The stimuli in this experiment were the same items as those used in Ostrand et al. (2016) Experiment 2 and further details can be found in that report, including the full list of stimuli. These words were minimal pairs, always differing in only the initial consonant. The incongruent words included audio ‘b’ + visual ‘d,’ audio ‘p’ + visual ‘t,’ audio ‘p’ + visual ‘k.’ audio ‘b’ + visual ‘g,’ and audio ‘m’ + visual ‘n’ pairings.

Procedure

Participants completed the identification task as part of a larger McGurk effect priming experiment (but not from Experiment 2 of Ostrand et al., 2016). Participants wore sound-insulated headphones while observing the speaker on a computer screen in front of them, and were instructed to watch and listen to each item carefully. They were shown each of the 36 incongruent prime stimuli, and used the keyboard to report the initial sound that they perceived at the start of each incongruent word.

Results

McGurk identifications

Identification responses to these incongruent stimuli were tabulated for proportion of visual- and auditory-consistent responses. The by-item average visual-consistent identification rate for these stimuli was 39.7% ($SE = 4.4\%$). The auditory-consistent identification rate for these stimuli was 35.8% ($SE = 3.0\%$), and not significantly different from the rate of visual identifications, $t(35) = 0.62$, $p = 0.270$, $d = 0.10$. Fourteen of the 36 incongruent primes (38.9%) were identified as the visual component more often than their auditory component. Auditory and visual identifications were of similar magnitude, with the auditory identifications slightly more common across items. The fact that the visual influence is substantially lower for the Ostrand et al. (2016) stimulus set, relative to that of the current study (68.4%), may be one reason why semantic priming appeared so different across studies. The data used for this analysis are available online (10.17605/OSF.IO/AD52R).

The rate of visual word identification for the stimuli of Experiment 1 (68.4%) was significantly greater than the visual word identification for the stimuli of Ostrand et al. (2016), $t(57) = 4.61$, $p < .001$ (two-tailed), $d = 1.23$. Similarly, the auditory consistent responses for the stimuli of Experiment 1 (19.5%) were significantly less than the auditory word identification for the stimuli of Ostrand et al. (2016), $t(57) = -3.88$, $p < .001$ (two-tailed), $d = -1.04$. Overall, these results indicate that the stimuli used in Experiment 1 did in fact produce more robust visual consistent McGurk effects than did the stimuli of Ostrand et al. (2016).

Discussion

The post hoc experiment indicates that the stimuli employed in Ostrand et al. (2016) did not support visual identification reliably more than they supported auditory consistent identification. This result could be related to why Ostrand et al. (2016) found that these stimuli were associated with semantic priming consistent with the auditory stimulus.

One limitation of the post-hoc experiment is that, in contrast to Experiment 1, one group of participants provided the identifications of the incongruent stimuli, and a different group of participants participated in the semantic priming task reported in Ostrand et al. (2016). Thus, it is possible that the participants in the post-hoc experiment experienced the McGurk effect for these stimuli differently than the participants in the priming experiment of Ostrand et al. (2016) did. However, prior work indicates that the McGurk effect is relatively stable within individual stimuli across participants (Basu Mallick et al., 2015), and thus the identification results from the post-hoc experiment are likely to be similar to those experienced by the participants in the priming task in Ostrand et al. (2016), even though the identification and priming results were derived from different sets of participants.

General discussion

The purpose of this research was to further investigate the time course of audiovisual integration in relation to lexical access. Ostrand et al. (2016) reported that with incongruent stimuli, semantic priming was consistent with the auditory signal of a multimodal input. This finding suggested that lexical access may commence prior to, or concurrent with, multisensory integration. This conclusion is at odds with prior multisensory speech data suggesting very early integration of the unimodal streams at the prephonemic level of linguistic processing (see Rosenblum, 2019, for a review). To better understand the relationship between word identification and semantic priming, we replicated that experiment with stimuli known to induce strong visually influenced responses, and collected free-response identification data on those incongruent stimuli from the same participants who completed the priming task.

With these changes, we found evidence that the semantic priming associated with audiovisual incongruent speech was consistent with the perceived (visual) component, rather than the nonperceived, auditory component. We further found that the degree of semantic priming from a particular incongruent prime's visual signal was correlated with the rate of visually consistent identifications for that incongruent stimulus. These results suggest that lexical access was performed using the perceived (and integrated) word. Importantly, these latter results suggest that lexical access may sometimes be based on the auditory component, specifically when the McGurk effect is less robust, and perception *is of* the auditory component.

The primary result of Ostrand et al. (2016) was a pattern of semantic priming consistent with the auditory word of the incongruent primes. This pattern at first might seem to contrast with the pattern observed in Experiment 1. However, the new identification results of the stimuli used by Ostrand et al. (2016) (reported here in the post hoc experiment) are informative in this regard. These results indicate that auditory identifications for the incongruent stimuli used by Ostrand et al. (2016) were substantially more common than were the auditory identifications for the stimuli of our Experiment 1. Thus, it is possible that Ostrand et al.' (2016) finding that the auditory channel of incongruent stimuli was often consistent with lexical access could be a result of participants often *perceiving* those incongruent stimuli as consistent with the auditory channel.

However, it is important to note that a number of factors distinguish Experiment 1 of the present work from that of Ostrand et al. (2016), and these differences could also account for the divergent results of these two experiments. Chief among these differences is the incongruent word-initial segment combinations used across experiments (b/v *vs.* b/d, b/g, p/t, p/k, m/n). Likewise, these experiments differed in the prime and target words tested, and thus the strength of the semantic relationship between them, as well as the talker used to create the stimuli. Any of these differences could have induced a different processing strategy such that the auditory *rather than* visual/perceptual component provided the basis for lexical access. It could be, for example, that only for b/v combinations does integration precede lexical access, and that for other combinations (e.g., b/d, b/g, p/t, p/k, m/n), lexical access occurs first. Future research can test this possibility.

Visual dominance versus fusion integration in the McGurk effect

It is worth noting that both the current study and the Ostrand et al. (2016) study used incongruent items which were created for *visual dominance*—in which the perception is that of the visual signal. In contrast, many incongruent stimuli produce a *fusion* identification, in which the perception is a combination of the auditory signal and the visual signal, and thus differs from both unimodal inputs. The choice to use visually dominant incongruent stimuli in the current study was based on an attempt to induce the strongest possible McGurk influence, as well as to limit the complexity of the counterbalancing and experimental design.

As the words used as target items in these experiments were semantically related to either the auditory or visual component of the incongruent prime stimulus, neither study was equipped to test the priming of incongruent words created via fusion, which could constrain the conclusions that can be made about audiovisual integration, as such, and its relation to lexical processing. However, visual-dominance McGurk effects are generally accepted as evidence of true multisensory integration, especially if participants are instructed to base responses on what they "hear," which indicates that the visual stimulus alters the auditory percept (e.g., Alsius et al., 2018; Rosenblum, 2019). For example, by asking participants to report what they "heard," the visual-consistent responses reported in Experiment 1 can be understood as reflecting instances in which the visual channel changed the perception of the auditory channel. In contrast, if the instructions had asked participants to report what they thought "the talker said," then visual-consistent responses could represent a mixture of participants who integrated the visual information with the auditory information (i.e., experienced the McGurk effect), and participants who heard the auditory channel correctly (i.e. did not experience the McGurk effect) but based their response on the visual signal (e.g., 'I heard … but I *saw* that the talker said…'). Moreover, applying the narrower fusion-only definition (i.e. when participants report hearing a word that is in neither the auditory or visual channel of the incongruent stimulus) misses relevant manifestations of the illusion (see Alsius et al. 2018).

Still, it is possible the semantic priming effects discussed here reflect semantic processing of the visual-alone, rather than integrated or auditory-alone, information. It may be the case that when presented with incongruent auditory and visual information, the comprehension system processes lexical information from a single modality before integration is complete. Ostrand et al. (2016) proposed this mechanism; though their theory suggested that lexical access occurs on the unimodal *auditory* stimulus before integration completes. However, the results of the ANCOVA (and its post hoc correlation) reported above argues against lexical access occurring on the unimodal visual-only signal. That analysis showed a positive relationship between the identification differential and priming differential of the stimuli: Stimuli which were more frequently perceived as the auditory channel (i.e., the primes that failed to support the McGurk effect) also show more priming for the auditory-related target word, compared with stimuli which were more frequently perceived as the visual channel. If lexical access occurred based on the unimodal visual information, as opposed to the integrated auditory + visual information, then there should be no by-item relationship between the identification results and priming results.

All the same, an important follow-up for future research is exploring lexical access using target words related and unrelated to a McGurk-fusion word, such as auditory 'bait' + visual 'gate' = perceived 'date' with semantically related target words of 'worm,' 'fence,' and 'time,' respectively. This experiment would allow researchers to dissociate semantic priming induced by the integrated auditory and visual information (e.g., 'date'; the "fusion" of auditory 'bait' + visual 'gate'), from that induced by either the unimodal auditory or the unimodal visual information.

Other evidence regarding the timing of lexical access and audiovisual integration

The primary motivation for Ostrand et al. (2016) was to investigate the relationship between the time course of audiovisual integration and lexical access. In light of the contrasting results of the present experiments and those of Ostrand et al. (2016), additional research into semantic priming and the McGurk effect will be necessary. However, it may be helpful to review some related findings.

An oft-cited example of later occurring multisensory integration is lexical context effects on the McGurk effect. Brancazio (2004) found that people were more likely to perceive the McGurk effect (i.e., integrate incongruent unimodal signals) when the two unimodal inputs integrated to form a word, as opposed to a nonword; in particular, McGurk effects were more common when the auditory signal was a nonword ('*besk*') compared with a real word ('*beg*'), and when the McGurk effect formed a real word ('*desk*') compared with a nonword ('*deg*'). Similarly, Experiment 1 of Ostrand et al. (2016) found that audiovisual primes with a real-word auditory signal induced the same priming effect regardless of whether they integrated to a McGurk nonword perception (auditory 'beef' + visual 'deef' = percept 'deef') or a congruent real-word perception (auditory 'beef' + visual 'beef' = percept 'beef'). However, these audiovisual primes with a real-word auditory signal elicited faster responses than those which had nonword auditory signals but were perceived as real words (auditory 'bamp' + visual 'damp' = percept 'damp'). These results could suggest that lexical access may proceed on the auditory signal alone if it is a real word, but wait for integration to complete if the auditory signal is a nonword. However, these lexical effects could also reflect interactions with processes associated with the McGurk effect that occur *after* multisensory integration such as post-integration phoneme categorization (see Brancazio, 2004, for a discussion; see also Alsius et al., 2018; Rosenblum, 2019, for discussions of inferences about multisensory integration from the McGurk effect).

There is also a literature concerning the phenomenon known as selective adaptation (Roberts & Summerfield, 1981; Saldaña & Rosenblum, 1994; Samuel & Lieblich, 2014) which supports later-occurring audiovisual speech integration. These studies report selective adaptation consistent with the (unperceived) auditory component of the incongruent stimulus. These studies suggest that selective adaptation is sensitive to preintegration speech information (Samuel & Lieblich, 2014), and supports the contention that multisensory integration occurs late in the lexical processing pipeline, in contrast to the results from the current work. However, Dorsi et al. (2021) recently reported selective adaptation effects which were consistent with a multisensory integration-supported phonemic restoration effect (i.e. visual speech + auditory noise results in participants hearing the noise as speech; see also Samuel, 1997; Warren, 1970), suggesting that selective adaptation may be sensitive to multisensory integration in some contexts. Further research should investigate the contrasting conclusions produced by these studies, as the temporal relationship between selective adaptation and multisensory integration remains an open question.

Conclusions

In conclusion, the results from the present experiment suggest that lexical processing is sensitive to the perceptual identification of a multisensory incongruent prime. Strong incongruent stimuli produced semantic priming consistent with the McGurk percept, and the consistency of that percept correlated with the degree of semantic priming from the McGurk percept. This conclusion supports the contention that multisensory integration occurs early in lexical processing. While it is possible that a similar effect helps to explain the results of Ostrand et al. (2016; Experiment 2), alternative explanations, such as phoneme-dependent processing strategies are also possible. Further work is still needed to fully understand the relationship between audiovisual integration identification and lexical access.

Appendix 1

Primes				
Audio	Visual	Congruency	Target	Type
Eat	Eat	Congruent	Pave	Word
Fork	Fork	Congruent	Rat	Word
Fail	Fail	Congruent	Owe	Word
Gas	Gas	Congruent	Road	Word
Friend	Friend	Congruent	Ring	Word
Fog	Fog	Congruent	Pull	Word
Fix	Fix	Congruent	Pierce	Word
Each	Each	Congruent	Nephew	Word
Nephew	Nephew	Congruent	Gurge	NonWord
Hurt	Hurt	Congruent	Flut	NonWord
Loose	Loose	Congruent	Glent	NonWord
Tease	Tease	Congruent	Kak	NonWord
Label	Label	Congruent	Glitch	NonWord
Throw	Throw	Congruent	Nenger	NonWord
Thick	Thick	Congruent	Nathing	NonWord
Wet	Wet	Congruent	Pring	NonWord
Music	Music	Congruent	Groke	NonWord
Job	Job	Congruent	Frah	NonWord
Later	Later	Congruent	Gleek	NonWord
Verb	Verb	Congruent	Prelith	NonWord
Meet	Meet	Congruent	Glob	NonWord
Morning	Morning	Congruent	Gorrer	NonWord
Night	Night	Congruent	Hace	NonWord
Lot	Lot	Congruent	Glith	NonWord
Tent	Tent	Congruent	Klesh	NonWord
Gums	Gums	Congruent	Fath	NonWord
Mouth	Mouth	Congruent	Gow	NonWord
Money	Money	Congruent	Gomp	NonWord
Road	Road	Congruent	Hahb	NonWord
Hog	Hog	Congruent	Flurve	NonWord
Head	Head	Congruent	Floth	NonWord
Shop	Shop	Congruent	Joat	NonWord
Bore	Gore	Incongruent	Sour	Word
Mine	Nine	Incongruent	Should	Word
Map	Nap	Incongruent	School	Word
Pod	Cod	Incongruent	Stomach	Word
Mail	Nail	Incongruent	Droke	NonWord
Pug	Tug	Incongruent	Coath	NonWord
Buy	Guy	Incongruent	Skeeling	NonWord
Might	Night	Incongruent	Geech	NonWord
Mice	Nice	Incongruent	Deesh	NonWord
Bait	Date	Incongruent	Cret	NonWord
Pad	Tad	Incongruent	Beeth	NonWord
Pie	Tie	Incongruent	Crub	NonWord
But	Gut	Incongruent	Deeth	NonWord
Bum	Gum	Incongruent	Bemp	NonWord
Part	Tart	Incongruent	Blent	NonWord
Pest	Test	Incongruent	Dreeve	NonWord

Appendix 1 displays the noncritical prime–target pairs used in the design of Experiment 1. Data from trials using these stimuli were not analyzed.

Acknowledgements This research was supported by National Science Foundation grant 1632530 to Lawrence D. Rosenblum.

Open practice statement The data for these experiments are available online (10.17605/OSF.IO/AD52R).

Authors' contributions Under the supervision L. Rosenblum, J. Dorsi designed the experiment and stimuli for Experiment 1 and supervised the data collected for that experiment. R. Ostrand designed the experiment and stimuli for the post hoc experiment and supervised the data collected for that experiment. J. Dorsi tabulated the data and ran the analysis for Experiment 1. R. Ostrand tabulated the data collected from the post hoc experiment and J. Dorsi ran the item level analysis on those tabulated data. J. Dorsi wrote the first draft of the manuscript and was responsible for incorporating substantial revisions based on comments and edits from L. Rosenblum and R. Ostrand.

Funding This research was supported by National Science Foundation (Grant 1632530) to Lawrence D. Rosenblum.

Data availability The data generated by the experiments reported here, and the critical stimuli used in the lexical decision task of Experiment 1 are available from the Open Science Framework (10.17605/OSF.IO/AD52R).

Code availability N/A.

Declarations

Ethics approval This research was conducted consistent with the principles of the Declaration of Helsinki. The Institutional Review Boards of the University of California, Riverside, and the University of California, San Diego approved of the procedures of Experiment 1 and the post hoc experiment respectively.

Consent to participate All participants who provided the data for the research reported here provided informed consent to participate prior to data collection.

Consent for publication All participants who provided the data for the research reported here provided informed consent for data collected from them to be published a scientific journal.

Conflicts of interest The authors are not aware of any conflicts of interest or competing interests related to the research reported here.

References

Alsius, A., Paré, M., & Munhall, K. G. (2018). Forty years after hearing lips and seeing voices: The McGurk effect revisited. *Multisensory Research, 31*(1/2), 111–144. https://doi.org/10.1163/22134808-00002565

Arnold, P., & Hill, F. (2001). Bisensory augmentation: A speechreading advantage when speech is clearly audible and intact. *British Journal of Psychology, 92*(2), 339–355. https://doi.org/10.1348/000712601162220

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*(2015). https://doi.org/10.1016/j.jml.2015.06.00

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. https://doi.org/10.1080/09541440601125623

Basu Mallick, D., Magnotti, J. F., & Beauchamp, M. S. (2015). Variability and stability in the McGurk effect: Contributions of participants, stimuli, time, and response type. *Psychonomic Bulletin and Review, 22*(5), 1299–1307. https://doi.org/10.3758/s13423-015-0817-4

Bebko, J. M., Schroeder, J. H., & Weiss, J. A. (2014). The McGurk effect in children with autism and asperger syndrome. *Autism Research, 7*(1), 50–59. https://doi.org/10.1002/aur.1343

Besle, J., Fischer, C., Bidet-Caulet, A., Lecaignard, F., Bertrand, O., & Giard, M.-H. (2008). Visual activation and audiovisual interactions in the auditory cortex during speech perception: Intracranial recordings in humans. *Journal of Neuroscience, 28*(52), 14301–14310. https://doi.org/10.1523/JNEUROSCI.2875-08.2008

Brancazio, L. (2004). Lexical influences in audiovisual speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 30*(3), 445–463. https://doi.org/10.1037/0096-1523.30.3.445

Brancazio, L., & Miller, J. L. (2005). Use of visual information in speech perception: Evidence for a visual rate effect both with and without a McGurk effect. *Perception & Psychophysics, 67*(5), 759–769. https://doi.org/10.3758/BF03193531

Brown, V. A., Hedayati, M., Zanger, A., Mayn, S., Ray, L., Dillman-Hasso, N., & Strand, J. F. (2018). What accounts for individual differences in susceptibility to the McGurk effect? *PLOS ONE, 13*(11), 1–20. https://doi.org/10.1371/journal.pone.0207160

Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C., McGuire, P. K., Woodruff, P. W., Iversen, S. D., & David, A. S. (1997). Activation of auditory cortex during silent lipreading. *Science, 276*(5312), 593–596. https://doi.org/10.1126/science.276.5312.593

Cho, H. C., & Abe, S. (2013). Is two-tailed testing for directional research hypotheses tests legitimate? *Journal of Business Research, 66*(9), 1261–1266.

Dias, J. W., & Rosenblum, L. D. (2011). Visual influences on interactive speech alignment. *Perception, 40*, 1457–1466. https://doi.org/10.1068/p7071

Dorsi, J. (2019). *Understanding how lexical and multisensory contexts support speech perception*. University of California, Riverside.

Dorsi, J., Rosenblum, L. D., Samuel, A. G., & Zadoorian, S. (2021). Selective adaptation in speech: Measuring the effects of visual and lexical contexts. *Journal of Experimental Psychology. Human Perception and Performance, 47*(8), 1023–1042. https://doi.org/10.1037/xhp0000769

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), 1–17. https://doi.org/10.1080/01690965.2012.701758

Fowler, C. A., Brown, J. M., & Mann, V. A. (2000). Contrast effects do not underlie effects of preceding liquids on stop-consonant identification by humans. *Journal of Experimental Psychology: Human Perception and Performance, 26*(3), 877–888. https://doi.org/10.1037/0096-1523.26.3.877

Goldinger, S. D. (1996). Auditory lexical decision. *Language and Cognitive Processes, 11*(6), 559–568. https://doi.org/10.1080/016909696386944

Green, K. P., & Kuhl, P. K. (1989). The role of visual information in the processing of place and manner features in speech perception. *Perception & Psychophysics, 45*(1), 34–42. https://doi.org/10.3758/BF03208030

Green, K. P., & Miller, J. L. (1985). On the role of visual rate information in phonetic perception. *Perception & Psychophysics, 38*(3), 269–276 http://www.ncbi.nlm.nih.gov/pubmed/4088819

Green K. P. & Norrix L. W. (2001). Perception of/r/and/l/in a stop cluster: Evidence of cross-modal context effects. *Journal of Experimental Psychology: Human Perception and Performance, 27*(1), 166.

Green, K. P., & Norrix, L. W. (1997). Acoustic cues to place of articulation and the McGurk effect: The role of release bursts, aspiration, and formant transitions. *Journal of Speech, Language and Hearing Reseach, 40*(3), 646–665. https://doi.org/10.1044/jslhr.4003.646

Kimmel, H. D. (1957). Three criteria for the use of one-tailed tests. *Psychological Bulletin, 54*(4), 351–353.

Kiss, G. R., Armstrong, C. A., & Milroy, R. (1972). *An associative thesaurus of English* (pp. 153–165). Medical Research Council, Speech and Communication Unit, University of Edinburgh.

MacDonald, J., & McGurk, H. (1978). Visual influences on speech perception processes. *Perception & Psychophysics, 24*(3), 253–257 http://www.ncbi.nlm.nih.gov/pubmed/704285

Magnotti, J. F., & Beauchamp, M. S. (2015). The noisy encoding of disparity model of the McGurk effect. *Psychonomic Bulletin & Review, 22*(3), 701–709. https://doi.org/10.3758/s13423-014-0722-2

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

Mitterer, H., & Reinisch, E. (2017). Visual speech influences speech perception immediately but not automatically. *Attention, Perception, & Psychophysics, 79*(2), 660–678. https://doi.org/10.3758/s13414-016-1249-6

Musacchia, G., Sams, M., Nicol, T., & Kraus, N. (2006). Seeing speech affects acoustic information processing in the human brainstem. *Experimental Brain Research, 168*(1/2), 1–10. https://doi.org/10.1007/s00221-005-0071-5

Namasivayam, A. K., Yiu, W., & Wong, S. (2015). Visual speech gestures modulate efferent auditory system. *Journal of Integrative Neuroscience, 14*(1), 73–83. https://doi.org/10.1142/S0219635215500016

Navarra, J., & Soto-Faraco, S. (2007). Hearing lips in a second language: Visual articulatory information enables the perception of second language sounds. *Psychological Research, 71*(1), 4–12. https://doi.org/10.1007/s00426-005-0031-5

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General, 106*(3), 226–254. https://doi.org/10.1037/0096-3445.106.3.226

Nelson, D., McEvoy, C. L., & Schreiber, T. A. (1998). The University of South Florida Word Association, Rhyme, and Word Fragment norms. *Behavior Research Methods, Instruments, & Computers, 36*(3) http://www.usf.edu/FreeAssociation/

Ostrand, R., Blumstein, S. E., & Morgan, J. L. (2011). When hearing lips and seeing voices becomes perceiving speech: Auditory-visual integration in lexical access. *Proceedings of the Annual Meeting of the Cognitive Science Society, 33*(33), 1376–1381. https://doi.org/10.1016/j.neuroimage.2010.12.063.Discrete

Ostrand, R., Blumstein, S. E., Ferreira, V. S., & Morgan, J. L. (2016). What you see isn't always what you get: Auditory word signals trump consciously perceived words in lexical access. *Cognition, 151*(2016), 96–107. https://doi.org/10.1016/j.cognition.2016.02.019

Pekkola, J., Ojanen, V., Autti, T., Jääskeläinen, I. P., Möttönen, R., Tarkiainen, A., & Sams, M. (2005). Primary auditory cortex activation by visual speech: An fMRI study at 3 T. *NeuroReport, 16*(2), 125–128. https://doi.org/10.1097/00001756-200502080-00010

Reisberg, D., McLean, J., & Goldfield, A. (1987). Easy to hear but hard to understand: A speechreading advantage with intact auditory stimuli. In B. Dodd & R. Campbell (Eds.), *Hearing by eye: The psychology of lip-reading* (pp. 97–113). Erlbaum.

Roberts, M., & Summerfield, Q. (1981). Audiovisual presentation demonstrates that selective adaptation in speech perception is purely auditory. *Perception & Psychophysics, 30*(4), 309–314. https://doi.org/10.3758/BF03206144

Rosenblum, L. D. (2019, August). Audiovisual speech perception and the McGurk effect. *Oxford Research Encyclopedia of Linguistics*, *August*, 1–28. https://doi.org/10.1093/acrefore/9780199384655.013.420

Rosenblum, L. D., & Saldaña, H. M. (1992). Discrimination tests of visually influenced syllables. *Perception & Psychophysics, 52*(4), 461–473. https://doi.org/10.3758/BF03206706

Rosenblum, L. D., Dias, J. W., & Dorsi, J. (2016a). The supramodal brain: Implications for auditory perception. *Journal of Cognitive Psychology, 5911*(1), 1–23. https://doi.org/10.1080/20445911.2016.1181691

Rosenblum, L. D., Dorsi, J., & Dias, J. W. (2016b). The impact and status of Carol Fowler's supramodal theory of multisensory speech perception. *Ecological Psychology, 28*(4), 262–294. https://doi.org/10.1080/10407413.2016.1230373

Ruxton, G. D., & Neuhäuser, M. (2010). When should we use one-tailed hypothesis testing? *Methods in Ecology and Evolution, 1*(2), 114–117.

Saldaña, H. M., & Rosenblum, L. D. (1994). Selective adaptation in speech perception using a compelling audiovisual adaptor. *The Journal of the Acoustical Society of America, 95*(6), 3658–3661. https://doi.org/10.1121/1.409935

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. https://doi.org/10.1016/S0167-6393(98)00051-X

Samuel, A. G. (1997). Lexical activation produces potent phonemic percepts. *Cognitive Psychology, 127*(2), 97–127.

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. https://doi.org/10.1037/a0036656

Sanchez, K., Miller, R. M., & Rosenblum, L. D. (2010). Visual influences on alignment. *Journal of Speech, Language, and Hearing Research, 53*, 262–272.

Stropahl, M., Schellhardt, S., & Debener, S. (2016). McGurk stimuli for the investigation of multisensory integration in cochlear implant users: The Oldenburg Audio Visual Speech Stimuli (OLAVS). *Psychonomic Bulletin & Review*, 1–10. https://doi.org/10.3758/s13423-016-1148-9

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America, 26*(2), 212–215.

Teinonen, T., Aslin, R. N., Alku, P., & Csibra, G. (2008). Visual speech contributes to phonetic learning in 6-month-old infants. *Cognition, 108*(3), 850–855. https://doi.org/10.1016/j.cognition.2008.05.009

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia, 45*(3), 598–607. https://doi.org/10.1016/j.neuropsychologia.2006.01.001

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science, 167*(3917), 392–393. https://doi.org/10.1126/science.167.3917.392

Windmann, S. (2004). Effects of sentence context and expectation on the McGurk illusion. *Journal of Memory and Language, 50*(2), 212–230. https://doi.org/10.1016/j.jml.2003.10.001