When Hearing Lips and Seeing Voices Becomes Perceiving Speech: Auditory-Visual Integration in Lexical Access

Rachel Ostrand (rostrand@cogsci.ucsd.edu)
University of California, San Diego, Department of Cognitive Science
9500 Gilman Drive, #0515, La Jolla, CA 92093-0515 USA

Sheila E. Blumstein (sheila_blumstein@brown.edu)
Brown University, Department of Cognitive, Linguistic, and Psychological Sciences
Box 1821, Providence, RI 02912 USA

James L. Morgan (james_morgan@brown.edu)
Brown University, Department of Cognitive, Linguistic, and Psychological Sciences
Box 1821, Providence, RI 02912 USA

Abstract

In the McGurk Effect, a visual stimulus can affect the perception of an auditory signal, suggesting integration of the auditory and visual streams. However, it is unclear when in speech processing this auditory-visual integration occurs. The present study used a semantic priming paradigm to investigate whether integration occurs before, during, or after access of the lexical-semantic network. Semantic associates of the un-integrated auditory signal were activated when the auditory stream was a word, while semantic associates of the integrated McGurk percept (a real word) were activated when the auditory signal was a nonword. These results suggest that the temporal relationship between lexical access and integration depends on the lexicality of the auditory stream.

Keywords: lexical access; McGurk Effect; auditory-visual integration; lexical-semantic network

Introduction

Speech comprehension is a complex, multi-staged process. Language input to the perceiver consists of information from several different sources which can augment the auditory speech stream, including visual information from the speaker's mouth and lip movements, knowledge about the speaker's accent and pronunciations, eye and head movements to highlight referents, and tone of voice and body language. While speech perception is most obviously driven by the auditory signal entering the listener's ears (Erber, 1975), visual information from a speaker's mouth and lip movements can affect and even significantly alter the perception of speech (Fort et al., 2010; Green, 1998; Summerfield, 1987), especially in noisy or degraded environments (Erber, 1975; Grant & Seitz, 2000; Sumby & Pollack, 1954). To be able to derive this processing contribution from visual information, the auditory and visual signals must be integrated into a single representation. The present work seeks to determine when such integration occurs during speech processing; in particular, whether it occurs before or after access to the lexical-semantic network.

McGurk and MacDonald (1976) first reported the McGurk Effect, in which incongruent audio and visual stimuli combine to induce in listeners the perception of a stimulus different than that of the actual sound input they have received. This effect is remarkable because of its illusory status – the listener perceives a token that is distinct from the sound signal, even with a perceptually good auditory exemplar. In this case, it is clear that the auditory and visual signals are integrated at some point during speech processing.

Theories of lexical retrieval in speech comprehension posit a mental lexicon as a repository of stored lexical items. This comprehension lexicon is an interconnected network of words, each containing the phonological, syntactic, and semantic information necessary for speech processing. To understand spoken speech, the incoming speech signal must activate its entry in the lexicon to retrieve the meaning of an input word (Aitchison, 2003; Collins & Loftus, 1975). This look-up process, using phonological input as a search key for its corresponding meaning, is known as lexical access. The present study investigates which components of the incoming speech stream influence this search process.

In the case of McGurk Effect stimuli, for which participants perceive a stimulus different from that presented by the auditory stream alone, the differing auditory and visual inputs were necessarily integrated at some point during speech processing. However, it is unclear whether this integration happens before, after, or coincidently with lexical access. That is, does the lexical representation which is ultimately activated for processing the speech input correspond to the auditory input alone, or to the combined auditory-visual percept, which may differ from that of the auditory signal? The study presented here investigates whether this combined percept is simply a perceptual illusion that fails to access the lexicon, or if the integrated percept is treated as input to the lexicon, thereby causing activation of its own semantic associates.

To create these integrated audiovisual-percepts, a video of a speaker mouthing an item is dubbed with an auditory track differing in the initial consonant's place of articulation.

Perceivers often perceive an item created in this manner not as the true auditory input, but as either a fusion of the auditory and visual signals or just the visual track alone. (For example, an auditory [ba] paired with a visual /ga/ often fuses to form the percept *da* while auditory [ba] paired with a visual /da/ may also be perceived as *da*.[1])

The phonological feature of place of articulation is more easily detected visually than are the features of manner and voicing (Binnie, Montgomery, & Jackson, 1974) and is also more susceptible to auditory noise interference (Miller & Nicely, 1955). Thus, the manner-place hypothesis for interpretation of incongruent audio-visual (AV) items suggests that the feature of place is contributed by the visual stream while the manner and voicing features are contributed by the auditory stream. The combination of these three features leads to an AV percept that can be distinct from that of the actual auditory signal (MacDonald & McGurk, 1978; Summerfield, 1987).

Visual Influences on Degraded Auditory Signals

Visual information can be particularly helpful for comprehending speech when the auditory signal is less than ideal. For example, as the signal-to-noise ratio (SNR) decreases, the improvement afforded by the addition of visual information strongly increases. Sumby and Pollack (1954) presented participants with congruent, bimodal videos and asked them to identify the words they detected. At extremely low signal-to-noise ratios (-30 dB), bimodal presentation increased lexical identification by 40 percentage points; at moderate SNRs, the additional visual information only increased identification by 20 percentage points, and at 0 SNR the increase in rate of identification was negligible. Similarly, combined auditory and visual speech presentation can withstand about a 5-10 dB worse SNR than can auditory-alone presentation while still maintaining a level of 80% correct identification (Erber, 1975). As the speech signal becomes less reliable, less information about the input can be gleaned from the auditory signal alone and thus the visual track has more of a chance to contribute. In line with this, Bastien-Toniazzo, Stroumza, & Cavé (2009) found higher incidence of McGurk Effect percepts in higher-noise environments, implying that with greater background noise comes a greater reliance on the visual signal, and thus a greater chance of integrating the two streams into a McGurk percept.

The same holds true for clear nonword auditory input. Brancazio (2004) found a strong lexical bias for incongruent McGurk videos, as the visual signal contributed more frequently when the auditory signal was a nonword than when it was a word. A nonword audio track is, in a way, comparable to a degraded stimulus – with no match in the lexicon, it could easily be the result of a hearing or speech segmentation error. Consequently, an accompanying visual signal may be treated as additional disambiguating information and thus taken more into account when interpreting the input of a nonword.

The auditory and visual streams of a bimodal stimulus enter the mind separately and independently and, at some point during lexical processing, are integrated to create a single, unified percept, as in the McGurk Effect. The present study investigates this integration process in relation to lexical access. There are three possible points at which the auditory and visual tracks could be integrated: before, after, or coincident with access to the lexicon. If AV-integration occurs before lexical processing, namely, early in the perceptual stages of speech comprehension, then the combined percept (not the auditory signal alone) should be treated as the input for the lexicon, and thus should access its own lexical-semantic entry and associates. This would also imply that AV-integration operates on purely bottom-up information: if the streams are integrated before they are looked up in the lexicon, integration cannot be dependent on the lexicality or non-lexicality of one or the other tracks.

An alternative possibility is that AV-integration occurs in post-lexical stages of processing. In this case, the two modalities would stay separated until one or both have been sent to the lexicon and either activated a match or not. Insofar as speech perception is fundamentally determined by the auditory signal, any priming effects should be those created by the auditory stimulus. Only later, after the lexicon has been accessed, would AV-integration take place, leading to the fused item that comprehenders perceive. As a result, the combined percept and the word or nonword it forms would have no contact with the lexicon and thus its lexicality would be irrelevant.

The final possibility is that AV-integration could occur during lexical access. In general, the two streams would enter the lexicon separately, where the auditory stream would likely be weighted more heavily as the primary modality of speech perception. If the auditory input is, for some reason, less than ideal – whether because it is degraded, or in noise, or not a real word – and thus cannot activate any lexical entry sufficiently to bring it to threshold, then any other available disambiguating information, including the visual signal, could be used to help resolve the identity of the input. As a result, if lexical access is delayed due to the poor quality of the auditory stimulus, AV-integration could take place during this time and thus affect the lexicon search outcome.

To compare these possibilities, two types of audio-visual incongruent prime stimuli were used: auditory-word/visual-nonword items, which, when integrated, lead to a nonword-percept, and auditory-nonword/visual-word items, which integrate to form a word-percept. If AV-integration occurs pre-lexically so that the two streams are combined early in processing, it is the combined McGurk percept that should

[1] Here, brackets ([X]) denote the auditory track of a stimulus; slashes (/X/) the visual track; and italics (*X*) the illusory percept resulting from the combination of the auditory and visual signals.

access the lexicon. In this case, the word-percept items should prime their associates but the nonword-percept items should not. Alternatively, if AV-integration happens later in the processing stream and is post-lexical, then priming should be dependent on the auditory input alone, and thus word-percept stimuli (with an auditory nonword) should not demonstrate priming while nonword-percept stimuli (with an auditory word) should.

Methods

Participants

Twenty-six Brown University undergraduates who were native English speakers and not fluent in any other languages participated in the experiment. Two subjects' data had to be discarded due to instrument malfunction. The remaining twenty-four participants ranged in age from 18 to 22 years, and all except one were right-handed. There were 13 males and 11 females in the group.

Materials

Each stimulus consisted of a bimodal prime, with either congruent or incongruent audio and visual streams, followed, after a 50 msec ISI, by an auditory-only target. Bimodal primes were defined as congruent if their audio and visual tracks came from the same utterance, and incongruent (McGurk) if they did not and thus the onset consonant presented in the signals did not match. Twenty-four of the incongruent bimodal primes were auditory-word/visual-nonword stimuli and 24 were auditory-nonword/visual-word stimuli. The congruent bimodal primes used the audio track from the analogous McGurk videos paired with their corresponding visual. For example, the McGurk video [beef]/deef/ had the corresponding congruent video [beef]/beef/. The initial consonant pairs used to create the McGurk videos were [auditory-/b/, visual-/d/], [auditory-/p/, visual-/t/], and [auditory-/m/, visual-/n/]. The intended McGurk percept formed by the incongruent videos was always the same as the visual track. As a result, for the incongruent stimuli, only one of the auditory and the McGurk percept was a real word, allowing for a clear picture of which signal was the cause of any observed priming effects.

Half of the audio-only targets were themselves evenly divided between semantically-related and unrelated words. The other half of the targets were nonwords. The semantically-related target words were chosen from the University of South Florida Free Association Norms database (Nelson, McEvoy, & Schreiber, 1998) and the Edinburgh Associative Thesaurus (Kiss et al., 1973). Where the associates provided by these two databases were nonexistent or the words deemed too long, associates were provided by lab members. Nonword targets were chosen from the ARC nonword database (Rastle, Harrington, & Coltheart, 2002) and were all one or two syllables long.

Design and Procedure

Participants were instructed to watch the videos and listen to the item that followed each. The task was to make a lexical decision on the second, auditory item by pressing either the "word" or "nonword" button on the button box placed in front of the subject. The assignment of word or nonword to each button was alternated between subjects. Participants were instructed to respond to the target word as quickly as possible. Stimuli were displayed in two blocks separated by a self-timed break.

Each participant saw the same prime video twice across the experiment, paired with either both nonword or both word targets. Importantly, each saw a McGurk and its corresponding congruent prime with the same two targets, so the reaction times could be directly compared by item within subject. Trials with the same prime were separated between blocks as were trials with the same target. Participants were given 7 practice trials at the start of the task which were not included in the final data analysis.

Results

Reaction times (RTs) were measured from the offset of the target item to reduce any potential effects of differences in the durations of the auditory targets, and were divided into two sets by McGurk percept lexicality. The RTs in each set were further separated into four categories based on the prime-target relationship: congruent-related, congruent-unrelated, incongruent-related, and incongruent-unrelated. Within each subject, any responses that were more than two standard deviations from the average RT of their category were removed, along with any items on which the participant's lexical decision response came before the onset of the target word or on which they made an incorrect response. The average latencies for the remaining items in each category were computed within-subject. Reaction time results for congruent-nonword/incongruent-word items (NW→W)[2] are presented in Figure 1 and congruent-word/incongruent-nonword items (W→NW) in Figure 2.

A 2 (congruency) x 2 (relatedness) repeated-measures ANOVA was conducted by participants separately for the NW→W items and for the W→NW items. For the NW→W items, there was a main effect of congruency (F(1, 23) = 6.197, p<.020), indicating that incongruent trials, which created a real word percept (e.g., [bamp]/damp/, perceived

[2] Congruent-nonword/incongruent word-percept items (e.g., [bamp]/bamp/ and [bamp]/damp/) will be referred to as NW→W. (This symbol will be used for both congruent and incongruent items.) This notation recalls the fact that in the incongruent stimulus, a nonword auditory stimulus *becomes* a word-percept through AV-integration. As the auditory tracks are the same for the congruent and incongruent stimuli of a pair, the lexicality of the congruent item is denoted by the first item of the pair (here, a nonword). Similarly, congruent-word/incongruent nonword-percept items (e.g., [beef]/beef/ and [beef]/deef/) will be denoted as W→NW.

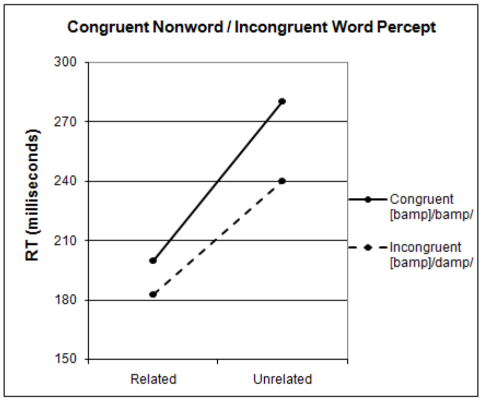

Figure 1: NW→W items

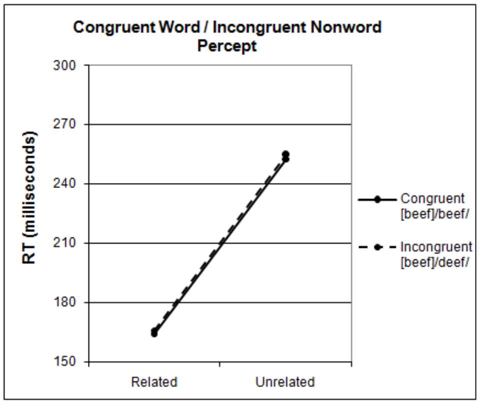

Figure 2: W→NW items

as *damp*), elicited faster response latencies to targets than did congruent nonword trials ([bamp]/bamp/, perceived as *bamp*). Additionally and importantly, there was a main effect of relatedness (F(1, 23) = 32.905, p<.000), demonstrating that priming indeed occurred, as related targets were responded to more quickly than were unrelated targets. There was no interaction (F(1, 23) = .744, p *ns*) between the factors.

The congruent word/incongruent nonword-percept (W→NW) stimuli behaved somewhat differently. As is evident from Figure 2, there was no main effect of congruency (F(1, 23) = .030, p *ns*) – the congruent and incongruent stimuli resulted in identical latencies for both related and unrelated prime-target pairs. There was, again, a significant main effect of relatedness (F(1, 23) = 41.413, p<.000). Unsurprisingly, there was no interaction (F(1, 23) = .002, p *ns*).

With these results in mind, a 2 (congruency) x 2 (relatedness) x 2 (percept lexicality) ANOVA was conducted including both stimulus types. There was a trend of a main effect of congruency, with RTs to incongruent-prime stimuli nearly significantly faster than to congruent-prime stimuli (F(1, 23) = 4.153, p<.053). There was a strong main effect of relatedness (F(1, 23) = 77.154, p<.000). There was additionally a strong main effect of percept lexicality (F(1, 23) = 7.528, p<.012), with RTs faster to W→NW stimuli than to NW→W stimuli. This result suggests that the auditory signal takes precedence over the visual: stimuli that formed real words without integration seem to have been activated more quickly than those that became lexical items only through the integration of the visual input. There was no interaction between any pair of two factors or of the three factors together.

All types of stimuli showed a significant effect of priming, as measured by strong main effects of relatedness in all comparisons.

Discussion

The goal of this study was to examine where in the processing stream auditory-visual integration occurs relative to lexical access. This question was investigated with regard to whether distinct auditory and visual tracks combine to form a McGurk Effect percept before or after the incoming signal is sent to the lexicon.

Three possibilities exist as to when in lexical processing auditory-visual integration could occur: before accessing the lexicon, after it, or simultaneously. The data support a hybrid account, in which AV-integration and lexical access occur in parallel and are inter-dependent.

The NW→W items demonstrated a strong effect of congruency: reaction times to targets paired with [bamp]/damp/ primes were faster than reaction times to targets paired with [bamp]/bamp/ primes. This makes a case for pre-lexical AV-integration. Both primes contained the same audio track, differing only by the fact that [bamp]/damp/ creates a real-word integrated percept (*damp*) while [bamp]/bamp/ remains a nonword (*bamp*). As reaction times following word primes are known to be faster than reaction times following nonword primes (e.g., Milberg et al., 1988), it seems to be the integrated, word-percept *damp* that accesses its lexical associates in the case of the incongruent NW→W stimulus, and thus AV-integration occurs before lexical access.

However, the W→NW items showed no effect of congruency: incongruent [beef]/deef/ and congruent [beef]/beef/ primes resulted in identical reaction times to both related and unrelated targets. This result suggests that AV-integration occurs **after** lexical access: as these items contained the same audio signal but differed in their visual signals, it appears that the auditory stimulus was driving the responses. Importantly, participants **did** integrate the auditory and visual information and perceived the combined McGurk percept in both the incongruent W→NW and NW→W cases – average goodness ratings as determined in a pilot experiment did not differ between these two groups of items.

Taken together, the results for the NW→W and the W→NW stimuli suggest that the influence of the integrated percept on lexical access depends on the lexical status of the auditory signal. When the auditory track is a word, integration takes place post-lexically so that the semantic associates of the audio signal become activated and primed – regardless of the congruency with the visual track. Conversely, if the audio is a nonword, then integration occurs before lexical access is complete, such that the combined, incongruent (word) percept results in significantly faster response times than does the congruent nonword stimulus. What's going on here?

In normal-hearing perceivers, speech comprehension is mainly driven by the auditory signal, as evidenced by the fact that auditory-only input is significantly more comprehensible than is visual-only input (Erber, 1975). In fact, listeners are adept at ignoring a visual speech signal when it could not have been generated by the same mechanism as the attended auditory speech stream and can use the auditory information exclusively if the visual signal is irrelevant or uninformative (Grant & Seitz, 2000). When presented with both auditory and visual information, the two streams, by virtue of the fact that they come from different modalities, enter the lexicon initially separate. While the two streams are still in the process of being integrated, a lexical search begins on the auditory signal, due to its privileged status in speech comprehension.

If the auditory signal is a word, it maps onto and activates that entry in the lexical-semantic network, thus priming its semantic associates. In this case, once a match has been found and a word has been selected, the integrated signal does not have a chance to influence lexical activation and selection. The actual integration of the auditory and visual information takes somewhat longer to complete than the spread of activation from the independent signals does, and thus occurs after the lexicon has already selected a word on the basis of the auditory signal alone.

The process begins in the same manner when the auditory signal is less than ideal – either because it is degraded, presented in a noisy environment, or is a nonword. Again, the auditory and visual signals enter the lexicon independently and not integrated, and the auditory signal spreads through the lexicon activating the sound structure and meaning it encodes. However, when the auditory input is a nonword, there is no matching lexical entry for it to activate. There is some partial activation of the nonword's phonological neighbors, but not enough to bring any individual word quickly to threshold. While this insufficient activation spreads through the lexicon, auditory-visual integration has a chance to complete. As no word has yet reached threshold and been selected, when the signal from the combined percept accesses the lexicon, it activates the integrated McGurk word. As a result, when the auditory signal cannot activate any one lexical entry enough to bring it to threshold **before** AV-integration takes place, and this integration results in a real word, it is the integrated percept's representation that is activated, leading to faster response times following incongruent word-percept primes than the corresponding congruent nonword primes.

An important component here is that response times to targets may be based on a different input stimulus than what the comprehender ultimately perceives. In general, with well-constructed McGurk stimuli, the comprehender should perceive a fused item, with the manner and voicing information contributed by the auditory track and the place of articulation supplied by the visual. However, in the case of an auditory-word incongruent (W→NW) stimulus, the auditory track activates its lexical representation and semantic associates in the lexicon before integration occurs. Thus the auditory signal determines the word actually activated in the lexicon while the combined audio and visual information determines the item the comprehender believes she has received.

This account of auditory-visual integration predicts that the congruent auditory-word items (e.g. [beef]/beef/) should activate their lexical entries faster than the incongruent word-percept items (e.g. [bamp]/damp/), which must wait for integration to take place before the lexical entry for the integrated percept can be activated. To test for this, a congruency x relatedness x percept lexicality ANOVA was conducted on the reaction time data. A strong main effect of lexicality emerged, with W→NW primes resulting in significantly faster reaction time latencies than NW→W primes. The W→NW items, composed from real-word auditory signals, could locate their input word in the lexicon pre-integration and thus spread activation to associated words *before* the AV-integration occurred, regardless of congruency. In contrast, incongruent NW→W items, while also resulting in a word percept, must wait for integration to occur before successfully finding a match in the lexicon, thus resulting in slower response times. The congruent NW→W items were simply nonwords and therefore result in slower RTs as well.

This account also explains why all four types of combinations showed equivalent related-unrelated priming, regardless of congruency or lexicality. The congruent and incongruent W→NW ([beef]/beef/ and [beef]/deef/) items should cause the same amount of priming as each other, as it is the identical auditory signal that is selected in the lexicon and thus the identical pattern of associates which is facilitated. For the incongruent NW→W ([bamp]/damp/) items, the integrated word percept activates its associates and thus results in the same amount of semantic facilitation as do the auditory-word items. Milberg and colleagues (1988) found a strong effect of mediated phonological-to-semantic priming. A nonword prime one phonological feature away from a real word elicited no difference in facilitation levels to a semantically-related target than did the real word prime itself. For example, *gat*, which differs from *cat* only by the initial consonant's place of articulation, primed *dog* to nearly the same extent that *cat* did. The congruent NW→W stimuli in this study (e.g. [bamp]/bamp/), while not forming real words, differed from real words by only a single feature; namely, the initial consonant's place of articulation, and thus, unsurprisingly, result in equivalent priming difference scores.

This account suggests some future directions for research. An important next step would be to repeat the study with incongruent prime stimuli consisting of both auditory and visual real words and a target semantically related to one of them (e.g., prime: [bait]/date/; target: [fish] or [time]). According to the present account, while the listener's perception may be that of the visual (i.e., integrated) signal, the semantic associates of the auditory signal should be primed. That is, we should observe significantly more facilitation for [bait]/date/–[fish] than for [bait]/date/–[time].

Additionally, varying the interstimulus interval between prime and target items may produce different patterns of results. Our account predicts that an extremely short ISI may leave no time for AV-integration of the prime before the target plays, and thus abolish the priming effect found for incongruent NW→W items. Conversely, a longer ISI might remove any reaction time differences between incongruent NW→W items and W→NW items as the two streams would have sufficient time to integrate before the lexical decision on the target had to be made.

In sum, the present study suggests that audiovisual integration occurs in parallel with lexical access. The auditory signal of a bimodal input is weighted more heavily as its activation moves through the lexicon, but if no lexical match is found by the time AV-integration occurs, the combined percept becomes the search item in the lexicon and can activate its semantic associates.

Acknowledgments

The authors thank Elena Tenenbaum and Naomi Feldman for providing many useful suggestions, John Mertus for his assistance with the technical components of the experiment, and Liz Chrastil for appearing as the model in the video stimuli. This work was conducted while the first author was at Brown University. It was funded in part by NIH grants DC00314 to SEB and HD32005 to JLM.

References

Aitchison, J. (2003). *Words in the Mind: An Introduction to the Mental Lexicon*. Malden, MA: Blackwell.

Bastien-Toniazzo, M., Stroumza, A., Cavé, C. (2009). Audio–visual perception and integration in developmental dyslexia: an exploratory study using the McGurk effect. *Current Psychology Letters [Online], 25*, URL: http://cpl.revues.org/index4928.html

Binnie, C.A., Montgomery, A.A., & Jackson, P.L. (1974). Auditory and Visual Contributions to the Perception of Consonants. *The Journal of Speech and Hearing Research*, *17*, 619-630.

Brancazio, L. (2004). Lexical influences in audiovisual speech perception. *Journal of Experimental Psychology, 30*, 445-463.

Erber, N. (1975). Auditory-Visual Perception of Speech. *Journal of Speech and Hearing Disorders, 40*, 481-492.

Fort, M., Spinelli, E., Savariaux, C., and Kandel, S. (2010). The word superiority effect in audiovisual speech perception. *Speech Communication*, *52*, 525-532.

Collins, A.M. & Loftus, E.F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82,* 407-428.

Ganong, W.F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, *6*, 110-125.

Grant, K.W., & Seitz, P.F. (2000). The use of visible speech cues for improving auditory detection of spoken sentences. *Journal of the Acoustical Society of America, 108*, 1197–1208.

Green, K.P. (1998). The use of auditory and visual information during phonetic processing: Implications for theories of speech perception. In R. Campbell, B. Dodd, & D. Burnham (Eds.), *Hearing by Eye II: Advances in the Psychology of Speechreading and Auditory-Visual Speech*. Hove, England: Psychology Press.

Kiss, G.R., Armstrong, C., Milroy, R., and Piper, J. (1973). An associative thesaurus of English and its computer analysis. In Aitken, A.J., Bailey, R.W. and Hamilton-Smith, N. (Eds.), *The Computer and Literary Studies*. Edinburgh: University Press.

MacDonald, J. & McGurk, H. (1978). Visual influences on speech perception processes. *Perception & Psychophysics*, *24*, 253-257.

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

Mertus J.A. (2008). BLISS: The Brown Lab Interactive Speech System. Brown University; Providence, RI.

Milberg, W., Blumstein, S., & Dworetzky, B. (1988). Phonological factors in lexical access: Evidence from an auditory lexical decision task. *Bulletin of the Psychonomic Society*, *26*, 305-308.

Miller, G.A. & Nicely, P.E. (1955) An analysis of perceptual confusions among some English consonants. *The Journal of the Acoustical Society of America*, *27*, 338-352.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). The University of South Florida word association, rhyme, and word fragment norms. http://www.usf.edu/FreeAssociation/.

Rastle, K., Harrington, J., & Coltheart, M. (2002). 358,534 nonwords: The ARC Nonword Database. *Quarterly Journal of Experimental Psychology*, *55A*, 1339-1362

Sumby, W.H. & Pollack, I. (1954). Visual Contribution to Speech Intelligibility in Noise. *Journal of the Acoustical Society of America, 26*, 212-215.

Summerfield, Q. (1987). Some preliminaries to a comprehensive account of audio-visual speech perception. In B. Dodd & R. Campbell (Eds.), *Hearing by Eye: The Psychology of Lip-Reading*. Hillsdale, NJ: Lawrence Erlbaum Associates Ltd.