DISCREPANT VISUAL SPEECH FACILITATES COVERT SELECTIVE LISTENING IN "COCKTAIL PARTY" CONDITIONS

JASON A. WILLIAMS

Summary.—The presence of congruent visual speech information facilitates the identification of auditory speech, while the addition of *incongruent* visual speech information often impairs accuracy. This latter arrangement occurs naturally when one is being directly addressed in conversation but listens to a different speaker. Under these conditions, performance may diminish since: (a) one is bereft of the facilitative effects of the corresponding lip motion and (b) one becomes subject to visual distortion by incongruent visual speech; by contrast, speech intelligibility may be improved due to (c) bimodal localization of the central unattended stimulus. Participants were exposed to centrally presented visual and auditory speech while attending to a peripheral speech stream. In some trials, the lip movements of the central visual stimulus matched the unattended speech stream; in others, the lip movements matched the attended peripheral speech. Accuracy for the peripheral stimulus was nearly one standard deviation greater with *incongruent* visual information, compared to the congruent condition which provided bimodal pattern recognition cues. Likely, the bimodal localization of the central stimulus further differentiated the stimuli and thus facilitated intelligibility. Results are discussed with regard to similar findings in an investigation of the ventriloquist effect, and the relative strength of localization and speech cues in covert listening.

Early examinations of speech intelligibility in noisy environments (e.g., Cherry, 1953; Cherry Taylor, 1954; Moray, 1959) constitute seminal work in cognitive psychology, revealing that individuals are able to explicitly report the content of only a single attended speech stream amongst simultaneously presented speech stimuli (the cocktail party effect). The results of these studies were highly influential in prompting the early-selection filter theory of attention (Broadbent, 1958); individuals can pattern- recognize only those stimuli attended to, and therefore attention may only be directed to pre-pattern recognition (i.e., physical-level) characteristics. Although Treisman (1964) and others subsequently revealed subtle behavioral effects for unattended stimuli, explicit attention remains a necessary condition for the recall of lengthy speech content (Pashler, 1999).

While a number of physical-level cues may be adequate to isolate stimuli in complex auditory environments (e.g., amplitude, Bregman, Abramson, Doehring, & Darwin, 1985; frequency, Bregman & Pinker, 1978; and timbre, Broadbent, 1952), much early research examined the importance of auditory localization in the differentiation of speech streams. Hirsch (1950) demonstrated that speech intelligibility of two-syllable words improved greatly during binaural compared to monaural listening, and in these former conditions was further facilitated by widening the distance between signal source and that of a white-noise generator. In an examination of protracted speech stimuli, participants presented with two simultaneous same-speaker orations from a single locus had great difficulty reporting either stream, but when these same orations were presented separately to each ear via headphones, they easily reported either (Cherry, 1953). Facilitative effects due to localization cues were also demonstrated with stimuli spatially separated by loudspeakers, with both two (Broadbent, 1954) and more (four and seven) simultaneous speakers (Pollack Pickett, 1958). More

recently Yost, Dye, and Sheft (1996) employed seven loudspeaker locations, reporting that the facilitative effect of spatially separating auditory stimuli rose in importance as the number of simultaneous voices increased from two to three. Overall, the facilitation of auditory localization has been shown to augment performance in numerous cocktail party environments (for reviews, see Yost, 1997; Bronkhorst, 2000).

A separate line of research has shown that visual stimuli can aid in auditory localization. Warren (1970) reported that pointing at visually occluded auditory targets became more accurate when participants had their eyes open. For Warren, two possibilities existed: Either (a) vision enabled a "structured visual environment" which facilitated performance, or (b) participants were making saccades to auditory targets and pointing where they were looking. When participants opened their eyes but were instructed to not make saccades, performance did not decrease, prompting Warren to favor the first hypothesis. Subsequent evidence that visual localization is superior to auditory localization (e.g., Blauert, 1983) also supports this interpretation. However, saccadic eye movements consist of two phases: a latency phase involving the calculation of the relative contractions of the three sets of ocular muscles and a ballistic phase where the movement is actually executed (Carpenter, 1988). It is possible to perform oculomotor programming without an actual eye movement, and this localization process may have accounted for increased accuracy in Warren's task. This possibility is strongly supported by Rorden and Driver (1999), who showed that the localization of auditory targets was facilitated by planning to look at the same hemi-field locations even though auditory stimuli had terminated prior to the occurrence of any eye motion. By either or both mechanisms, visual information has been shown to aid auditory localization.

Given that performance in "cocktail party" environments improves as localization is facilitated, it should be the case that visual localization cues aid performance under these conditions. Direct evidence was provided by Driver (1996) in an investigation of the ventriloquist effect, the phenomenon whereby auditory stimuli are localized toward the presence of a viable visual source for the auditory stimuli (Thurlow & Jack, 1973). In Experiment 1, Driver (1996) presented individuals with simultaneous auditory stimuli (paired sequences of three-words) from a single location peripheral to fixation. Simultaneously, a computer monitor displayed the lip movements of the attended auditory sequence either proximal to the auditory source or in the contralateral hemi-field. This second condition, in which the video was further displaced from the location of the auditory stimuli, resulted in increases in identification accuracy of nearly 2070. Presumably the auditory stimulus congruent with the displaced visual speech became further localized toward the visually presented face, perceptually increasing the separation of the stimuli and aiding discrimination. With regard to the present study, if a perceptual illusion that localizes speech can facilitate performance, veridical visual localization information should as well.

Consider an arrangement that occurs regularly in everyday life. Covert attention refers to the ability to look in one direction while attending another (Posner, 1980). With auditory stimuli, this often occurs when individuals are feigning listening to someone addressing them and attending to a conversation elsewhere. To successfully comprehend the peripheral auditory stimulus, one must ignore not only the auditory speech from the unattended speaker's oration, but also the corresponding visual speech information. In question is the perceptual effect of this unattended visual information, of which there are conflicting possibilities: (a) the addition of a conficting stimulus may serve as an additional distractor that interferes with performance or (b) akin to Driver's (1996) findings with the ventriloquist effect, additional visual information may aid in

localizing an auditory stream, and result in increased speech intelligibility. One clear difference from Experiment 1 of that study, however, is that Driver provided participants with visual information that matched an attended auditory channel. In covert attention, however, the argument is that visual information *incongruent* with the attended speech stream may facilitate performance.

METHOD

Participants

Twenty-eight men and 20 women (M_{Age} =19.7 yr., *SD* =1.3, range =18- 23) participated as a partial requirement for Introductory Psychology courses. All participants were treated according to the university's Human Subjects Committee's conditions for experimental approval, including informed consent and full debriefing, and were sequentially assigned to one of 24 counter-balanced conditions in the order that they arrived at the experiment.

Materials and Procedure

Each participant sat 45 cm in front of a 32.5 cm x 24.4 cm computer monitor, to which was affixed a 5.1-cm diameter loudspeaker on the lower frame. Central speech stimuli originated from a location as near as possible to the mouth area of faces that were subsequently displayed on the video screen. For half of the participants, an identical loudspeaker was affixed to the (quiet, but not sound-proof) laboratory wall at the same height peripherally 1.2 m to the right of the central display; for the other half, the peripheral loudspeaker was situated similarly on the left side.

Full-face visual stimuli were recorded by digital video (MPEG—2 media format) at the time of auditory recording. Auditory stimuli consisted of recordings of individuals uttering short sentences generated according to the following syntactical structure: "The *noun past-tense verb* the *adjective noun*" (e.g., "The fog obscured the hidden army"), avoiding predictable semantic regularities (e.g., "The farmer plowed..."). One-hundred-twenty such sentences were digitally recorded, matched for approximate utterance length, and stored electronically as separate audio channels in the MPEG—2 file with 5-sec. intervals of silence separating the pairs. During the experiment, a Dell Optiplex 790 PC computer played the video in full- screen mode, and the 60 stimulus pairs were presented with one speech stream centrally, and the other peripherally at approximately 60dB. In all trials, participants were asked to report the content of the peripheral auditory stream while simultaneously viewing a central display which varied across trial blocks. Participants recorded their responses immediately after each trial on answer sheets that contained starter sentences containing the two constant 'the's, and blanks for the four target words ("The ____ ____ the ____ ____").

Each participant evaluated 12 auditory pairs in four differing visual conditions: (a) the *Visual Speech Present with Modal Discrepancy* (VSP—MD) in which the monitor displayed the face of the person mouthing the centrally located auditory speech, akin to naturalistic settings. However, since participants were always listening to the peripheral stimulus, the visual speech information contrasted with the attended spoken sentences, and thus was bimodally discrepant; (b) a *Visual Speech Absent with Modal Discrepancy* (VSA—MD), in which a static picture of the central speaker's face was substituted for the moving video[3] in order to examine performance in the absence of visual speech information and still mirror the prior condition as closely as possible. Because the viewed face was mismatched to

[3]Simply seeing a face may affect performance, and therefore this seemed preferable to presenting a blank screen.

the person speaking the attended auditory stimuli, stimuli were again bimodally discrepant; (c) a Visual Speech Present with Modal Congruency (VSP—MC), in which dynamic visual speech now corresponded to the attended auditory peripheral stimulus. This arrangement entailed that the visual and auditory cues from the central location were in conflict; however, since the participants were listening peripherally, the visual stimulus was congruent with the attended auditory speech content, and thus the speech content was bimodally congruent; (d) a Visual Speech Absent with Modal Congruency (VSA—MC), in which the visual display consisted of a static face of the individual who was speaking the attended peripheral speech stream, and thus mismatched the central auditory stimuli. Since Kamachi, Hill, Lander, and Vatikiotis-Bateson (2003) reported that individuals cannot accurately match voices to same-sex still photographs, from the participants' perspective this condition was likely perceptually identical to the VSA—MD condition (an effect replicated in the present study and discussed in the Results section).

To control for any interstimulus differences in difficulty (e.g., articulation, slight volume changes, word frequency, salience, etc.), every particular auditory pair was evaluated with equal frequency in each of the four visual conditions. To accomplish this, all auditory stimuli were presented in an identical sequence to each participant, but the order of visual conditions was varied between participants to ensure that each auditory pair was presented an equal number of times in each visual condition. Specifically, trials were divided into five blocks of 12 pairs each, with the first block discarded as practice (this initial block employed female speakers; Blocks 1 and 3 were males, and 2 and 4 were females). Given that there might be interaction effects between the four levels (e.g., some conditions may induce more practice or fatigue than others and affect performance on subsequent blocks), it seemed best to control for any such effects by fully counterbalancing the sequences of visual conditions. Therefore, against the fixed auditory presentation, equal numbers of participants were exposed to the four differing blocks of visual stimuli in each of the 24 possible sequences, with one-half receiving the peripheral stimulus situated to the left, and one-half to the right (*n* 48). After the completion of all data collection, success in word identification (exact matches-only, spelling mistakes ignored) was tallied and summed for each participant, and a score out of 48 possible (four target words in each of 12 trials per block) constituted the measure of speech intelligibility for each experimental condition.

RESULTS

None of the participants volunteered the existence of vision or hearing issues, or reported difficulties seeing or hearing the stimuli presented; all were able to hold normal conversations with the experimenters prior to and after data collection. Given the within-group design employed, individual differences in visual and auditory acuity were balanced across all levels of the four conditions. No individual score was further than two standard deviations from the mean within any condition, and no data were excluded from analysis.

Initial omnibus testing was accomplished via a repeated-measures analysis of variance (ANOVA) across the four visual conditions, and was statistically significant ($F_{3,45}$ 9.6, $p <$ .001). Means and standard deviations are presented in Table 1. A subsequent analysis compared the two manipulations without visual speech information, and as expected there was virtually no difference between modality congruent ($M_{VSA–MC}$ 24.06, SD = 7.7) and modality discrepant ($M_{VSA–MD}$ 24.08, SD = 8.5) conditions (Tukey test, df = 47, p =.99), replicating the findings of Kamachi, *et al*. (2003). Since these were unfamiliar, participants had no idea whether a particular unmoving face was congruent or discrepant with any particular voice. Given the

previous research indicating that the control conditions were perceptually identical, and that the observed means were virtually indistinguishable, these were averaged together to arrive at a single *Visual Speech Absent* (VSA) measure for each participant.

A second repeated-measure ANOVA compared the Visual Speech Present Modal Discrepant (VSP–MD) condition, the Visual Speech Present Modal Congruent (VSP–MC) condition, and the merged VSA condition. Results were statistically significant ($F_{2,46}$ 14.7, $p < .001$) and *post hoc* tests (Tukey test, $df = 47$) revealed that all conditions were significantly different from each other. Compared to participants' performance with no visual-speech information (M_{VSA} = 24.07, SD = 6.3), the addition of the modality-discrepant moving face (akin to covert listening) significantly ($p < .001$, d =.55) increased accurate identification ($M_{VSP–MD}$ = 27.5, SD = 6.1) in spite of the fact that the extra visual information conflicted with the veridical response. In fact, when visual speech matched the peripherally attended voice, performance (M_{VSR-MC} = 21.7, SD = 7.6) significantly decreased (p =.041, d =.34) compared to no visual information (M_{VSA} = 24.07, SD = 6.3). Finally, when directly comparing the moving conditions, bimodally discrepant information (M_{VSP-MD} = 27.5, SD = 6.1) was shown to be much superior ($p < .001$, d =.85) to bimodally congruent information (M_{VSP-MC}= 21.7, SD = 7.6).

TABLE 1
DESCRIPTIVE STATISTICS FOR ALL CONDITIONS

	No Visual Speech		Visual Speech	
	M	*SD*	*M*	*SD*
Face and voice match modally	24.06	7.7	21.7	7.6
	(VSA-MC)		(VSP-MC)	
Face and voice mismatch modally	24.08	8.5	27.5	6.3
	(VSA-MC)		(VSP-MD)	

Note.—48 correct responses were possible: 12 trials per block, with 4 targets per trial. VSA– MC Visual Speech Absent, Modality Congruent; VSP–MD = Visual Speech Present, Modality Congruent; VSA–MD Visual Speech Absent, Modality Discrepant; VSP–MD =Visual Speech Present, Modality Discrepant.

DISCUSSION

The present examination of covert listening to peripheral speech in a "cocktail party" environment produced three findings of note. Firstly, for bimodal discrepant visual speech facilitation, compared to an absence of visual speech, participants were significantly more accurate in identifying the words presented peripherally when the central visual display of the speaker's mouth/lip movements matched the auditory stimulus (sounds and words in sentences) presented centrally. This occurred despite the conflict of the additional visual information with accurate responses. Secondly, in bimodal congruent visual speech hindrance, compared to no visual speech information, the addition of centrally presented visual speech that matched the attended peripheral speech reduced performance despite the considerable evidence that shows that congruent visual speech information aids intelligibility in noisy environments (e.g., Sumby & Pollack, 1954). Thirdly, in the condition without voice/face matching, replicating the findings of Kamachi, *et al.* (2003), individuals were unable to match unfamiliar voices to still images of faces; as the implications of that finding were discussed in the study, they are not discussed here.

Discrepant Visual Speech Facilitation (VSP–MD Condition)

The current results are among few in the literature to demonstrate that discrepant visual speech can aid in intelligibility. In Exp. 3 of Driver (1996), same-speaker target and distractor speech were simultaneously presented from two extremely close locations, proximal to a video display on which was presented either no-visual speech or visual speech congruent with the distractor stimulus. The latter condition resulted in significant increases in the intelligibility of targets, and Driver argued that the addition of discrepant visual information further segmented the distractor sounds from the target sounds via cross-modal integration. The present experiment expands the scope of this phenomenon to ecologically valid conditions that employed different orators and spatially separated auditory stimuli.

Facilitation effects for visually discrepant speech are unexpected when one considers previous research on mismatched bimodal speech presented in isolation. When auditory and visual stimuli conflict, for example the presentation of the auditory phoneme /ba/ with a visually presented /ga/, the two are combined to arrive at the most likely perceptual candidate, and may result in a perception different from either of the sources, in this particular case /da/ (McGurk MacDonald, 1976). The phenomenon, termed the "McGurk effect," which alters the perceived place of articulation, is not limited to consonants, but also occurs with vowels (Summerfield McGrath, 1984) and whole words (Dekle, Fowler, Funnell, 1992).

The VSP—MD condition of the present experiment would seem to constitute an environment of central visual speech, which would be discrepant to peripherally attended auditory stimuli, and therefore accuracy, accordingly, should have declined; instead, discrepant speech significantly facilitated intelligibility. As bimodal interference likely still occurred, the most probable explanation for the observed result is that this negative effect was offset by a separate factor that significantly facilitated intelligibility, likely bimodal localization of the central stimulus. Theoretical support for this interpretation is discussed in the Introduction, and also reinforced by the finding of location discrepant speech hindrance in the VSP—MC condition.

Location Discrepant Visual Speech Hindrance (VSP–MC Condition)

Given the considerable evidence that lip-motion aids intelligibility in noisy environments (for a comprehensive review see Summerfield, 1992), the addition of visual speech information consistent with the attended auditory stimulus would seem to facilitate intelligibility. However, performance actually decreased when modality-congruent visual information was introduced compared to VSA conditions. It is unclear whether bimodal speech facilitation occurred with spatially separated stimuli (an issue discussed further below), but if it did, it was outweighed by a much larger adverse effect, the most probable candidate being the severe disruption of localization cues between modalities.

This interpretation is supported by an analysis of Driver's (1996) experiments. When two auditory stimuli were presented from a single location (Exp. 3), discrepant visual speech further segmented the stimuli and facilitated intelligibility of the target speech. However, when localization cues existed to distinguish the stimuli, i.e., when the auditory stimuli were laterally separated in Exp. 2, the addition of discrepant visual speech located between them decreased intelligibility. In this arrangement, the ventriloquist effect served to perceptually locate the distractor speech location more centrally, reducing perceived spatial separation and thus performance. Any segmentation facilitation that may have occurred was therefore trumped by the attenuation of localization cues.

Conclusions, Limitations, and Future Directions

Across Exps. 2 and 3, Driver (1996) held visual information constant, manipulated the spatial separation of auditory stimuli, and found that discrepant visual speech aided speech intelligibility *unless* it disrupted localization cues. The current investigation held the spatial separation of auditory stimuli constant, manipulated the visual speech information (congruent vs. discrepant), and arrived at a similar result. When localization was facilitated and speech information disrupted (VSP—MD trials), intelligibility increased. However, when speech information was facilitated and localization disrupted (VSP—MC trials), intelligibility decreased. Given the large (*d* .85) intelligibility advantage in the former condition compared directly to the latter, during covert listening with two speakers, bimodal localization is significantly more important than the presence of bimodal speech cues.

Beyond this assessment, a more quantitative approach to the relative magnitude of these cues could be addressed by a future study that employed multiple speaker locations, manipulated sound sources across a range of eccentricities, and had participants assess central as well as peripheral stimuli. For example, if the two auditory streams were minimally separated and near the visual source, by comparing visual and still conditions one could measure the facilitative effect of visual speech information for that particular set of stimuli. One could then have participants report the central speech stream and vary the presentation of the distractor speech at varying eccentricities, with and without visual speech information, and isolate the relative contribution of visual speech as localization cues become more salient. In addition, the effect of temporal contiguity could be investigated with this design by varying the synchrony of the auditory and visual speech.

Despite the present investigation, Driver's (1996) experiments, and the numerous investigations of both bimodal speech facilitation and the McGurk effect (McGurk MacDonald, 1976), one central question remains unanswered: does bimodal speech integration, either facilitative or intrusive, occur between a fixated visual stimulus and an attended auditory stimulus, or a location-consistent stimulus? Additional possibilities are that synthesis occurs if either situation applies, or only when both conditions are present. In classic investigations of the McGurk effect and bimodal facilitation, visual speech is paired to auditory stimuli both by attention to that stimulus, and by originating from a common location. Experiments employing location-congruent visual and auditory stimuli cannot differentiate between any of these possibilities.

Experimental arrangements with multiple speech stimuli also have difficulty addressing the issue. For example, in Exp. 2 of Driver (1996)

with spatially separated stimuli, it is unclear whether the (facilitative) segmentation effect found in Exp. 3 still occurred but was masked by the attenuation of localization cues, or if segmentation requires spatial contiguity and simply did not occur. The current study is likewise limited in regards to the McGurk effect and cross-modal interference in general. Compared to VSA, the addition of discrepant speech (VSP—MD) facilitated intelligibility, presumably due to the addition of bimodal localization cues of the central stimulus. What remains unknown is whether McGurk interference occurred between the central stimulus and spatially separated (but attended) peripheral stimulus, and was masked by localization facilitation, or if the effect requires spatial contiguity and did not occur. The same limitation applies to an analysis of bimodal facilitation in the VSP— MC condition; perhaps spatially separated facilitation did occur but was masked by the disruption of localization cues.

An investigation of this problem may require an examination of implicit processing of the unattended central stimuli. Despite a lack of explicit recall for unattended speech, Swinney (1979) and others have demonstrated semantic priming to such stimuli, and it may be possible to exploit this phenomenon to address bimodal speech integration. Consider the naturally occurring covert listening condition in the present study (VSP—MC). By employing a lexical decision task or other priming-detection paradigms, one could presumably determine whether the McGurk synthesis occurred with unattended location-congruent stimuli. For example, suppose an auditory stimulus "ball' is presented in the unattended central channel simultaneous with an auditory stimulus "gall' in the attended peripheral channel. In VSA conditions, presumably both auditory stimuli would be primed. However, when visual speech is presented from the central location, attention-based and location-based accounts offer differing predictions regarding the unattended channel. If the synthesis is attention-based, only "ball' should be primed (although perhaps more strongly compared to VSA given bimodal integration). However, if the synthesis is location based, sub-threshold perception of the unattended channel should be to the synthesized composite "dall.' A subsequent lexical decision task should be able to assess whether this occurs, answering whether attention is necessary for the McGurk effect, or whether spatial contiguity is sufficient.

In conclusion, the results suggest that during selective covert listening in multiple speech environments, localization processes are particularly significant. In naturalistic environments, a plethora of evidence for facilitation effects exists for additional spatially and content-congruent visual speech information; the present evidence shows that adding visual speech stimuli also aids in the intelligibility of an attended peripheral

stream. Whether this facilitation would extend to environments containing more than two speakers is unknown, but is likely given that the addition of modally congruent visual speech information has been shown to increase in importance as the number of voices increases (Rudmann, Mc-Carley, & Kramer, 2003). Finally, future variations of this paradigm may shed light on attention-based versus location-based integration of various bimodal speech stimuli.

REFERENCES

BLAUERT, J. (1983) *Spatial hearing.* Cambridge, MA: MIT Press.

BREGMAN, A. S., ABRAMSON, J., DOEHRING, P., & DARWIN, C. J. (1985) Spectral integration based upon common amplitude modulation. *Perception & Psychophysics,* 37, 483- 493.

BREGMAN, A. S., & PINKER, S. (1978) Auditory streaming and the building of timbre. *Canadian Journal of Psychology*, 32, 19-31.

BROADBENT, D. E. (1952) Listening to one of two synchronous messages. *Journal of Experimental Psychology*, 44, 51-55.

BROADBENT, D. E. (1954) The role of auditory localization in attention and memory span. *Journal of Experimental Psychology*, 47, 191-196.

BROADBENT, D. E. (1958) *Perception and communication.* London: Pergamon Press.

BRONKHORST, A. W. (2000) The cocktail party phenomenon: a review of research on speech intelligibility in multiple-talker conditions. *Acta Acoustica*, 86, 117-128.

CARPENTER, R. M. S. (1988) *Movements of the eyes.* London: Pion Limited.

CHERRY, E. C. (1953) Some experiments on the recognition of speech, with one and two ears. *Journal of the Acoustical Society of America,* 25, 554-559.

CHERRY, E. C., & TAYLOR, W. K. (1954) Some further experiments upon the recognition of speech, with one and two ears. *Journal of the Acoustical Society of America*, 26, 975-979.

DEKLE, D. J., FOWLER, C. A., & FUNELL, M. G. (1992) Audiovisual integration in perception of real words. *Perception & Psychophysics*, 51(4), 355-362.

DRIVER, J. (1996) Enhancement of selective listening by illusory mislocation of speech sounds due to lip-reading. *Nature*, 381, 66-68.

HIRSCH, I. J. (1950) The relation between localization and intelligibility. *Journal of the Acoustical Society of America, 22*, 196-200.

KAMACHI, M., HILL, H., LANDER, K., & VATIKIOTIS-BATESON, E. (2003) "Putting the face to the voice": matching identity across modality. *Current Biology*, 13(19), 1709-1714.

MCGURK, H., & MACDONALD, J. (1976) Hearing lips and seeing voices: a new illusion. *Nature*, 264, 746-748.

MORAY, N. (1959) Attention in dichotic listening: affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology*, 11, 56-60.

PASHLER, H. E. (1999) *The psychology of attention*. Cambridge, MA: MIT Press.

POLLACK, I., & PICKETT, J. M. (1958) Stereophonic listening and speech intelligibility against voice babble. *Journal of the Acoustical Society of America*, 30, 131-133.

POSNER, M. I. (1980) Orientation of attention: the VIIth Sir Frederic Bartlett lecture. *Quarterly Journal of Experimental Psychology, 32A*, 3-25.

RORDEN, C., & DRIVER, J. (1999) Does auditory attention shift in the direction of an upcoming saccade? *Neuropsychologia, 37*, 357-377.

RUDMANN, D. S., MCCARLEY, J. S., & KRAMER, A. F. (2003) Bimodal displays improve speech comprehension in environments with multiple speakers. *Human Factors*, 45(2), 329-336.

SUMBY, W. H., & POLLACK, I. (1954) Visual contributions to speech intelligibility in noise. *Journal of the Acoustical Society of America*, 26, 212-215.

SUMMERFIELD, Q. (1992) Lipreading and audio-visual speech perception. P*hilosophical Transactions: Biological Sciences*, 335(1273*)*, 71-78.

SUMMERFIELD, Q., & MCGRATH, M. (1984) Detection and resolution of audio-visual incompatibility in the perception of vowels. *Quarterly Journal of Experimental Psychology*, 36A, 51-74.

SWINNEY, D. A. (1979) Lexical access during sentence comprehension: reconsideration of some context effects. *Journal of Verbal Learning and Verbal Behavior*, 18, 645-659.

THURLOW, W. R., & JACK, C. E. (1973) A study of certain determinants of the "ventriloquism effect." *Perceptual and Motor Skills*, 36, 1171-1184.

TREISMAN, A. M. (1964) The effect of irrelevant material on the efficiency of selective listening. *American Journal of Psychology*, 77, 533-546.

WARREN, D. H. (1970) Intermodality interactions in spatial localization. *Cognitive Psychology*, 1, 114-133.

YOST, W. A. (1997) The cocktail party problem: forty years later. In R. H. Gilkey & T. R. Anderson (Eds.), *Binaural and spatial hearing in real and virtual environments.* Mahwah, NJ: Erlbaum. Pp. 329-348.

YOST, W. A., DYE, R. H., JR., & SHEFT, S. (1996) A simulated "cocktail party" with up to three sound sources. *Perception & Psychophysics*, 58(7), 1026-1036.